UNIFORM CRITICAL VALUES FOR LIKELIHOOD RATIO TESTS IN BOUNDARY PROBLEMS

GIUSEPPE CAVALIERE[a], ADAM MCCLOSKEY[b], RASMUS S. PEDERSEN[c] AND ANDERS RAHBEK[c]

July 25, 2025

ABSTRACT

Limit distributions of likelihood ratio statistics are well-known to be discontinuous in the presence of nuisance parameters at the boundary of the parameter space, which lead to size distortions when standard critical values are used for testing. In this paper, we propose a new and simple way of constructing critical values that yields uniformly correct asymptotic size, regardless of whether nuisance parameters are at, near or far from the boundary of the parameter space. Importantly, the proposed critical values are trivial to compute and at the same time provide powerful tests in most settings. In comparison to existing size-correction methods, the new approach exploits the monotonicity of the two components of the limiting distribution of the likelihood ratio statistic, in conjunction with rectangular confidence sets for the nuisance parameters, to gain computational tractability.

Uniform validity is established for likelihood ratio tests based on the new critical values, and we provide illustrations of their construction in two key examples: (i) testing a coefficient of interest in the classical linear regression model with non-negativity constraints on control coefficients, and, (ii) testing for the presence of exogenous variables in autoregressive conditional

[a]Department of Economics, University of Bologna, Italy and Department of Economics, University of Exeter, UK.
[b]Department of Economics, University of Colorado, USA.
[c]Department of Economics, University of Copenhagen, Denmark.
Correspondence to: Adam McCloskey, email adam.mccloskey@colorado.edu.
We thank Pascal Lavergne and Frank Kleibergen for comments and suggestions. We also thank participants at the 2022 CIREQ Montreal Econometrics Conference, the 2023 Robust Econometric Methods in Financial Econometrics Workshop in Copenhagen, the 2023 ICEEE Conference in Cagliari, and the 2024 Toulouse Financial Econometrics Conference for comments.
McCloskey acknowledges support from the National Science Foundation under Grant SES-2341730. A. Rahbek and G. Cavaliere gratefully acknowledge support from the Independent Research Fund Denmark (DFF Grant 7015-00028) and the Italian Ministry of University and Research (PRIN 2020 Grant 2020B2AKFW).

heteroskedastic models (ARCH) with exogenous regressors. Simulations confirm that the tests have desirable size and power properties. A brief empirical illustration demonstrates the usefulness of our proposed test in relation to testing for spill-overs and ARCH(-X) effects.

KEYWORDS: Likelihood ratio tests; Parameters on the boundary; Uniform inference.

INTRODUCTION

AS IS NOW WELL-KNOWN, the asymptotic distribution of a likelihood ratio (LR) statistic is discontinuous under the null hypothesis being tested at points for which a vector nuisance parameter lies on the boundary of its parameter space. This feature complicates the proper formation of critical values (CVs) that control the asymptotic size of the LR test uniformly over the parameter space. In this setting, we propose a simple and computationally-tractable method of CV formation (Algorithm 1) that controls the asymptotic size of a test using the standard LR statistic. The correct uniform asymptotic size of the resulting test (Theorem 3.3) ensures that the test also has good size properties in finite samples whether the nuisance parameter is at, near or far away from the boundary of its parameter space.

Although a few recent advances in the literature have now produced hypothesis tests that are uniformly valid over the nuisance parameter space, our focus in this paper is to introduce a hypothesis test that is simple to compute even when the dimension of the vector of nuisance parameter is not small, without sacrificing much power. To attain this goal, we continue to use a standard LR statistic because it is one of the most widely-used tests in statistics and is efficient in finite samples under correct specification when nuisance parameters are far from the boundary. However, the standard χ^2-based CVs for LR statistics are invalid when nuisance parameters are on or near the boundary. In contrast, the CVs we propose are based upon the asymptotic distributions of this test statistic that arise under certain parameter sequences that drift toward the boundary of the parameter space.

In order to establish uniform asymptotic size control, we derive the null asymptotic distribution of LR statistics under a comprehensive class of parameter sequences that drift the nuisance parameter vector toward the boundary at any sample size-dependent rate. Under a particularly important rate for establishing asymptotic size, the asymptotic distribution of the LR statistic depends upon a nuisance parameter vector that cannot be estimated consistently but can still be "estimated" by an asymptotically Gaussian random vector centered at its true value. Our key insight is that confidence bounds for this latter "estimator" can be combined with monotonicity properties inherent to the two components of the asymptotic distribution of the LR statistic to yield CVs that can be computed via straightforward Monte Carlo simulation. In order to uniformly control asymptotic size, we make use of a standard Bonferroni correction to account for the randomness involved in the "estimation" of the nuisance parameter.

The main theoretical result of our paper (Theorem 3.3) proves that our algorithm for CV formation (Algorithm 1) controls axymptotic size uniformly in a general framework encompassing numerous and varied applications. Examples include tests of a one-dimensional sub-vector of the mean in the multivariate Gaussian location model with a restricted mean vector, tests of a regression coefficient in the linear regression model when some coefficients have a known sign and tests of parameters in random coefficients models such as the workhorse empirical industrial organization model of Berry, Levinsohn and Pakes (1995). We verify that the conditions of our main theoretical result hold in applications of tests on regression coefficients and specification tests in ARCH-type models, the latter being a pervasive example of tests suffering from boundary problems in the recent literature on conditional volatility models, see, e.g., Francq and Zakoïan (2009), Cavaliere, Nielsen and Rahbek (2017), Cavaliere, Nielsen, Pedersen and Rahbek (2022) and Cavaliere, Perera and Rahbek (2024). We contribute to this literature by providing the first hypothesis test we are aware of with proven uniform asymptotic validity for specification testing for the presence of exogenous variables in ARCH-type models.

Several papers in the literature derive asymptotic distributions of estimators and test statistics at the boundary of the parameter space; see, e.g., Self and Liang (1987), Shapiro (1989), Geyer (1994), Silvapulle and Silvapulle (1995) and Andrews (2001). More recently, Cavaliere et al. (2022) propose a bootstrap-based CV construction that implicitly uses an estimator to switch between the quantiles of the asymptotic distribution at and far away from the boundary in large samples. However, using CVs that correspond to distributions at the boundary and/or far away from the boundary does not ensure uniform asymptotic size control since null distributions can have larger quantiles in intermediate ranges near the boundary of the parameter space.

This latter fact has been recognized in the literature, leading to inference methods that control asymptotic null rejection probabilities uniformly across all parameter sequences that may drift toward the boundary of the parameter space. In particular, Andrews and Guggenberger (2009) propose CVs in a general testing framework that are asymptotically equivalent to least-favorable CVs which find the maximal quantile of a test statistic's null asymptotic distribution over the nuisance parameter that cannot be consistently estimated. Recent alternative approaches allowing for quite general constraints, including Hong and Li (2020) and Li (2025), advocate similar least favorable-type CVs that are derived from sophisticated bootstrap implementations. Recognizing the conservative nature of tests using these CVs, and its negative consequence on power, McCloskey (2017) proposes CVs that do not maximize these quantiles over the entire nuisance parameter space but rather a first stage confidence set for the nuisance parameter; see also Berger and Boos (1994) and Silvapulle (1996) for parametric finite-sample versions of this approach. In the context of boundary problems, Mitchell, Allman and Rhodes (2019) apply McCloskey's (2017) approach to CV formation for LR statistics although, unlike the current paper, they focus on problems that may also feature singularities (see Drton, 2009). Fan and Shi (2023) also take this approach to CV formation for Wald and Quasi-Likelihood Ratio (QLR) statistics in more complicated boundary problems that require an initial first step of identifying an implicit nuisance parameter.

Although McCloskey's (2017) method tends to produce better power properties than the least favorable approach of Andrews and Guggenberger (2009) and related papers, both methods can become computationally intractable when the nuisance parameter is not low-dimensional because they generally require the optimization of a function whose values must be simulated at each parameter value. The approach we propose in this paper is similar in spirit to the approach of McCloskey (2017), but by exploiting the monotonicity properties of the two components of the LR statistic, remains computationally tractable in the presence of nuisance parameters that are not low-dimensional because it does not require optimization of a function calculated via simulation.

Finally, Ketz (2018) and Ketz and McCloskey (2023) propose computationally tractable and uniformly valid inference approaches to inference when parameters may be on or near the boundary of the parameter space. Although these approaches are different from that taken in this paper, one commonality these two works share with our approach here is that they use one-step estimators, discussed e.g., in Newey and McFadden (1994), to attain asymptotically Gaussian estimation as input to form valid inference regardless of where the true parameter lies in relation to the boundary of their parameter space.

Structure of the paper. The remainder of the paper is organized as follows. Section 1 provides basic intuition for our proposed method. Section 2 introduces the general setting and assumptions as well as the ongoing examples on hypothesis testing in a linear regression model with positivity constrained coefficients and specification testing in ARCH models. Section 3 contains our proposed algorithm for critical value construction and proofs the validity of the critical values. Section 4 considers the performance of our proposed method by means of Monte Carlo simulations, and Section 5 contains a small empirical illustration in relation to specification testing in ARCH models for various stock market indices. All proofs and additional simulations are collected in the supplemental appendix.

Notation. The following notation is used throughout the paper. The set of positive definite matrices with eigenvalues bounded below by some $\kappa > 0$, and above by κ^{-1} is denoted by Υ. For a square $m \times m$ matrix A, let $\text{diag}(A)$ a diagonal matrix with the same diagonal of A. Moreover, let $\text{diagv}(A)$ denote the m-dimensional column vector with the diagonal elements of A as entries. $\mathbb{I}()$ is the indicator function. With A a positive definite matrix and x a vector, $\|x\|_A = \sqrt{x'Ax}$.

1 Main ideas in a simple Gaussian model

To obtain the basic intuition underlying our proposed CV construction, consider the case of a scalar parameter of interest γ and a scalar nuisance parameter β, where $\gamma \geq 0$ and $b \geq 0$. We consider testing $\mathsf{H}_0 : \gamma = 0$, against the alternative, $\mathsf{H}_1 : \gamma > 0$. For simplicity, we frame this in terms of the a bivariate normal model for $Y \in \mathbb{R}^2$, where

$$Y = (Y_\gamma, Y_b)' \sim \mathrm{N}\left(\lambda, \Sigma\right),$$

with $\lambda = (\gamma, b)'$ and Σ a known positive definite covariance matrix. In standard settings, this model corresponds to a limiting experiment. For testing H_0, consider the standard likelihood ratio (LR) statistic as defined by

$$\mathsf{LR} = \inf_{\lambda \in \{0\} \times [0,\infty]} (\lambda - Y)'\Sigma^{-1}(\lambda - Y) - \inf_{\lambda \in [0,\infty] \times [0,\infty]} (\lambda - Y)'\Sigma^{-1}(\lambda - Y).$$

Under H_0, by standard arguments

$$\mathsf{LR} \sim \inf_{\lambda \in \{0\} \times [-b,\infty]} Q(\lambda) - \inf_{\lambda \in [0,\infty] \times [-b,\infty]} Q(\lambda) \equiv \mathcal{L}(b, b), \tag{1.1}$$

for $Q(\lambda) = (\lambda - Z)'\Sigma^{-1}(\lambda - Z)$ with $Z \sim \mathrm{N}(0, \Sigma)$.

The difficulty with controlling the size of a test using the LR statistic in this setting comes down to the fact that β is unknown and therefore the null distribution $\mathcal{L}(b, b)$ in (1.1) is unknown. Since the null distribution is unknown, it is not obvious how to define a critical value (CV) such that the test has size no greater than a nominal level $\alpha \in (0, 1)$. A naive approach which sets the CV to the $1 - \alpha$ quantile of (1.1) at some given value of b can lead to size distortions. For example, the standard CV is equal to the $1 - \alpha$ quantile of (1.1) for $b = \infty$, corresponding to a $\max\{\mathrm{N}(0, 1), 0\}^2$-distribution. However, a test using this CV will over-reject if the $1-\alpha$ quantile of the distribution in (1.1) is larger for some value of $b < \infty$. To illustrate this, the figure below contains the 95-percentile of $\mathcal{L}(b, b)$ for the case where Σ is a correlation matrix with correlation parameter ρ.

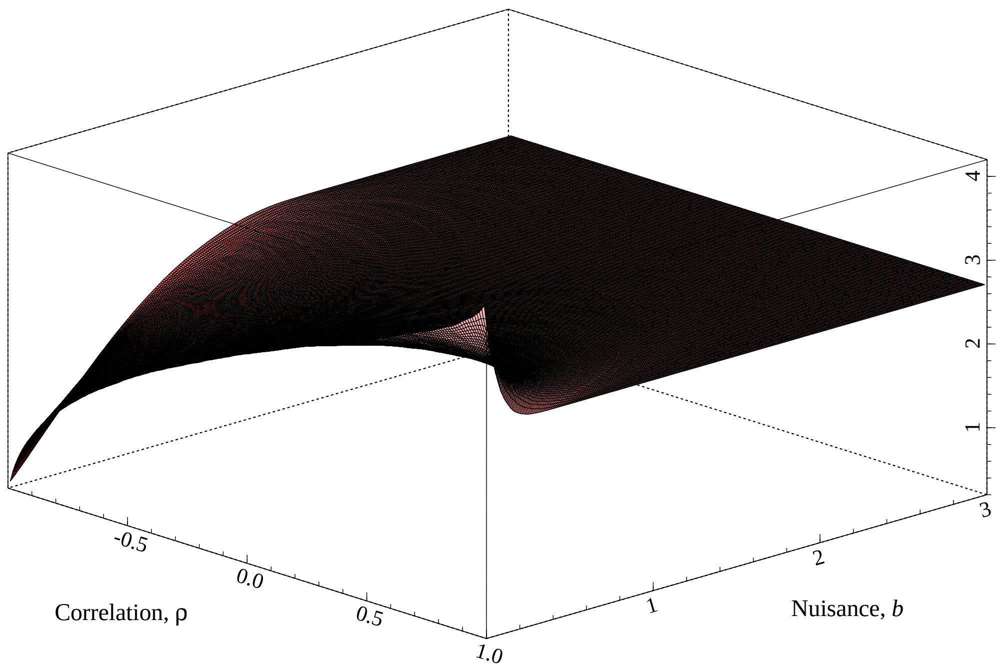

Figure: Simulated 95%-quantiles of $\mathcal{L}(b,b)$ as a function of $\rho \in (-1,1)$ and $b \geq 0$.

A solution, referred to as the "least favorable approach" of Andrews and Guggenberger (2009), is based on using the largest the $1-\alpha$ quantile of (1.1) across all possible values of $b \in [0,\infty]$ and therefore has correct size. Although not a major concern in the present example for which b is one-dimensional, the least favorable approach becomes computationally intractable as the dimension of the nuisance parameter grows beyond two or so. Furthermore, this approach can be very conservative, leading to poor power when the true value of b is not close to the value used to compute the CV.

Our alternative CV is constructed as follows. Define

$$\mathcal{L}(x,y) = \inf_{\lambda \in \{0\} \times [-x,\infty]} Q(\lambda) - \inf_{\lambda \in [0,\infty] \times [-y,\infty]} Q(\lambda),$$

where $\mathcal{L}(x,y)$ is stochastically decreasing in x, while increasing in y. Furthermore, let $\mathsf{CV}_q(x,y)$ denote the qth quantile of $\mathcal{L}(x,y)$ and note: (i) under H_0, $\mathsf{LR} \sim \mathcal{L}(b,b)$, and, (ii) for any values $b_L < b_U$ with $b \in (b_L, b_U)$, $\mathsf{CV}_q(b,b) \leq \mathsf{CV}_q(b_L,b_U)$. We use these observations to choose b_L and b_U in a judicious data-dependent manner to feasibly control the size of the LR test. Specifically, choose $\eta \in (0,\alpha)$ and let $z_{1-\eta/2}$ denote the $(1-\eta/2)^{th}$

quantile of $\mathrm{N}(0,1)$. The with Σ_{22} second diagonal entry of Σ and

$$b_L = Y_\beta - \sqrt{\Sigma_{22}} z_{1-\eta/2}, \quad b_U = Y_\beta + \sqrt{\Sigma_{22}} z_{1-\eta/2}, \tag{1.2}$$

it holds that $\mathbb{P}(b \in (b_L, b_u)) = 1 - \eta$. Thus, our proposal is to use $\mathsf{CV}_\alpha = \mathsf{CV}_{1-\alpha+\eta}(b_L, b_U)$ as a CV. To see how it controls size, note that for any $b \geq 0$ the probability of rejecting under H_0 can be bounded above using Bonferroni's inequality:

$$\mathbb{P}\left(\mathcal{L}(b,b) \geq \mathsf{CV}_\alpha\right) \leq \mathbb{P}\left(\mathsf{CV}_{1-\alpha+\eta}(b,b) \geq \mathsf{CV}_\alpha\right) + \mathbb{P}\left(\mathcal{L}(b,b) \geq \mathsf{CV}_{1-\alpha+\eta}(b,b)\right).$$

The sum of the two terms on the right hand side of the above bound is less than α since

$$\mathbb{P}\left(\mathsf{CV}_{1-\alpha+\eta}(b,b) \geq \mathsf{CV}_\alpha\right) = 1 - \mathbb{P}\left(\mathsf{CV}_{1-\alpha+\eta}(b,b) < \mathsf{CV}_\alpha\right) \leq 1 - \mathbb{P}(b \in (b_L, b_U)) = \eta,$$

and $\mathbb{P}(\mathcal{L}(b,b) \geq \mathsf{CV}_{1-\alpha+\eta}(b,b)) = \alpha - \eta$.

As we demonstrate in Theorem 3.3, CV_α can be generalized to control size as a CV for LR statistics when (a) the parameter space for γ is not constrained, (b) the dimension of β exceeds one, and, (c) (the estimators) Y_γ and Y_b are not normally distributed in finite samples. For (a), the minimization space $[0, \infty]$ in the expressions for LR and $\mathcal{L}(b,b)$ are modified to properly reflect the parameter space for γ under H_1. For (b), in order to use the multivariate version of the monotonicity property (ii), b_L and b_U are replaced by multivariate counterparts derived from rectangular confidence sets for the mean of a multivariate normal distribution. For (iii), the CVs described in Section 3 below do not require normally distributed estimators but rather estimators that are asymptotically Gaussian. We describe how to obtain such estimators in the presence of boundary constraints via a one-step Newton-Raphson iteration below.

2 General setup and framework

2.1 Setup

For a given set of observations $\{W_t\}_{t=1}^n$, consider the (log-likelihood) objective function,

$$L_n(\theta) = f_n(\{W_t\}_{t=1}^n; \theta), \tag{2.1}$$

in terms of the parameter vector $\theta = (\gamma', \beta', \delta')' \in \Theta = \Theta_\gamma \times \Theta_\beta \times \Theta_\delta$, where $\gamma \in \mathbb{R}^{d_\gamma}$ is the parameter (vector) of interest, while $\beta \in \mathbb{R}^{d_\beta}$ and $\delta \in \mathbb{R}^{d_\delta}$ are nuisance parameters. Specifically, we throughout assume that the true value of δ be an interior point of Θ_δ, while the true value of β is allowed to be in the interior, or at (near) the boundary of Θ_β. Furthermore, for γ we assume s of the components are in the "interior", while the remaining $d_\gamma - s$ are allowed to be on the "boundary", $0 \leq s \leq d_\gamma$. Formally (without loss of generality), $\Theta_\gamma = [\gamma_L, \gamma_U]^s \times [0, \gamma_U]^{d_\gamma - s}$, $\Theta_\beta = [0, \beta_U]^{d_\beta}$, with $-\infty \leq \gamma_L < 0 < \gamma_U \leq \infty$, $0 < \beta_U \leq \infty$, and $\Theta_\delta \subset (-\infty, \infty)^{d_\delta}$ is a compact subset. Finally, as in Andrews and Cheng (2012), we introduce an additional parameter ϕ used here to capture any features of the distribution of the data $\{W_t\}$ not explicit from the formulation in (2.1). This way, $\psi = (\theta, \phi)$ 'completely determines the distribution of the data'. The parameter space is $\Psi = \{\psi = (\theta, \phi) : \theta \in \Theta, \phi \in \Phi(\theta)\}$ where $\Phi(\theta) \subset \Phi$ with Φ a compact metric space which induces weak convergence of the bivariate distributions (W_t, W_{t+s}), all $t, s \geq 1$.

We are interested in testing the hypothesis

$$\mathsf{H}_0 : \gamma = \gamma_0$$

by using the (quasi-) LR statistic. That is, with the unrestricted $\hat{\theta}_n$ and restricted $\tilde{\theta}_n$ estimators as defined by

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta) \quad \text{and} \quad \tilde{\theta}_n = \arg\max_{\theta \in \Theta_{\mathsf{H}_0}} L_n(\theta),$$

where $\Theta_{\mathsf{H}_0} = \{\theta = (\gamma', \beta', \delta')' \in \Theta : \gamma = \gamma_0\}$, the statistic is given by,

$$\mathsf{LR}_n = 2(L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)). \tag{2.2}$$

To provide context and intuition, we include as running examples linear regression models with sign-restricted coefficients as well as a time series ARCH model with explanatory covariates.

Running example: Regression. With $W_t = (y_t, x_t')'$ consider the regression equation,

$$y_t = \theta' x_t + \varepsilon_t, \text{ for } t = 1, 2, \ldots, n, \tag{2.3}$$

with $x_t = (x_{1,t}, x_{2,t}')'$, $\theta = (\gamma, \beta')' \in \Theta = \Theta_\gamma \times \Theta_\beta = [0, \infty]^{1+d_\beta}$. We are interested in testing $\mathsf{H}_0 : \gamma = 0$, which we emphasize is a non-standard testing problem as some of entries of β may be zero-valued such that β is a boundary point of Θ_β.

Given a sample $\{W_t\}_{t=1}^n$, the least-squares objective function is given by

$$L_n(\theta) = -\tfrac{1}{2}\sum_{t=1}^{n}(y_t - x_t'\theta)^2 = -\tfrac{n}{2}\left(S_{yy} + \theta' S_{xx}\theta - 2\theta' S_{xy}\right),$$

with $S_{yy} = n^{-1}\sum_{t=1}^n y_t^2$, $S_{xx} = n^{-1}\sum_{t=1}^n x_t x_t'$, and $S_{xy} = n^{-1}\sum_{t=1}^n x_t y_t$. With the ordinary least-squares estimator given by $\hat{\theta}_{LS} = S_{xx}^{-1}S_{xy}$, it follows that the unrestricted estimator is given by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta}(\theta - \hat{\theta}_{LS})' S_{xx}(\theta - \hat{\theta}_{LS}),$$

see Lemma A.4 in the Appendix, while the restricted estimator is given by

$$\tilde{\theta}_n = \arg\min_{\theta\in\Theta_{\mathsf{H}_0}}(\theta - \hat{\theta}_{LS})' S_{xx}(\theta - \hat{\theta}_{LS}),$$

where $\Theta_{\mathsf{H}_0} = \{\theta = (\gamma, \beta')' \in \Theta : \gamma = 0\}$, and, finally, the LR statistic is given by (2.2).

Finally, we note that, in this example, the additional parameter ϕ determines the joint distribution of x_t and ε_t. $\square$

Running example: ARCH. For the second example, we consider hypothesis testing in ARCH models augmented with non-negative explanatory (or exogenous, X) covariates. Let $y_t \in \mathbb{R}$ be given by

$$y_t = \sigma_t \varepsilon_t, \quad t \in \mathbb{Z}, \tag{2.4}$$

$$\sigma_t^2 = \theta' F_{t-1} \tag{2.5}$$

where $F_t = (X_t', Y_t', 1)'$, with $X_t = (x_{1,t}, \ldots, x_{p,t})'$ and $Y_t = (y_t^2, \ldots y_{t-q+1}^2)'$. We are interested in testing whether the covariates are needed for the conditional variance σ_t^2. With θ partitioned as $\theta = (\gamma', \beta', \delta)'$, let $F_t = (Y_{\gamma,t}', Y_{\beta,t}', 1)'$, write (2.5) as a function of θ as

$$\sigma_t^2(\theta) = \theta' F_{t-1} = \delta + \beta' Y_{\beta,t-1} + \gamma' Y_{\gamma,t-1},$$

with $\theta \in \Theta = \Theta_\gamma \times \Theta_\beta \times \Theta_\delta$, $\Theta_\gamma = [0, \gamma_U]^p$, $\Theta_\beta = [0, \gamma_U]^q$ and $\Theta_\delta = [\delta_L, \delta_U]$, $\delta_L > 0$. As before, the objective is to test the hypothesis $\mathsf{H}_0 : \gamma = 0$, which is a non-standard testing problem as some of the β's can be on the boundary.

Given a sample $\{W_t\}_{t=1}^n = \{(y_t^2, F_{t-1}')\}_{t=1}^n$, the Gaussian quasi-log-likelihood function is (up to a constant) given by

$$L_n(\theta) = \sum_{t=1}^n l_t(\theta), \qquad l_t(\theta) = -\frac{1}{2}\left(\log\sigma_t^2(\theta) + \frac{y_t^2}{\sigma_t^2(\theta)}\right). \tag{2.6}$$

The QMLE $\hat\theta_n$ ($\tilde\theta_n$) is any maximizer of $L_n(\theta)$ over Θ ($\Theta_{\mathsf{H}_0} = \{\theta \in \Theta : \gamma = 0\}$), and

$$\mathsf{LR}_n(\mathsf{H}_0) = 2\left[L_n(\hat\theta_n) - L_n(\tilde\theta_n)\right] \tag{2.7}$$

is the LR statistic.

Finally, in this example the additional parameter $\phi = \phi(\theta)$ determines the joint distribution of F_{t-1} and ε_t. □

2.2 Drifting sequences and assumptions

In order to establish uniform validity of our proposed critical values in the possible presence of nuisance parameters near or on the boundary, we study the limiting behavior of the LR statistic in (2.2) under drifting sequences of parameters $\psi_n = (\theta_n, \phi_n) \to \psi_0 = (\theta_0, \phi_0)$. Specifically, we let $\theta_n = (\gamma_0', \beta_n', \delta_n')'$ denote a sequence of parameters in Θ_{H_0}, that is, a sequence along which H_0 holds and which satisfy $\theta_n \to \theta_0 = (\gamma_0', \beta_0', \delta_0')'$ as $n \to \infty$. In order to allow β to be either an interior point or near/at the boundary, in addition to $\beta_n = \left(\beta_{n,1}, \ldots, \beta_{n,d_\beta}\right)' \to \beta_0$ we require that $\sqrt{n}\beta_n \to b = (b_1, b_2, \ldots, b_{d_\beta})' \in [0,\infty]^{d_\beta}$, as $n \to \infty$. This includes sequences of true parameters converging to an interior point – e.g., $b_i = \infty$ – or to a boundary point at the standard $\sqrt{n}$ rate – e.g., $b_i \in [0, \infty)$. Moreover, parameters converging to zero at a rate slower (faster) than $\sqrt{n}$ corresponds to, e.g., $b_i = \infty$ ($b_i = 0$). Henceforth, the distribution of the (stationary) random vectors $\{W_t : t \geq 1\}$ is determined by the true parameter ψ; expectations, variances and probabilities computed under ψ are denoted as $\mathbb{E}_\psi$, $\mathbb{V}_\psi$ and $\mathbb{P}_\psi$, respectively.

In order to state the limiting distribution of LR_n, we make the following assumptions which are similar to Andrews (2001); see also Ketz (2018) and Fan and Shi (2023). These assumptions are stated for any drifting sequence $\psi_n \in \Psi$ satisfying

$$\psi_n \to \psi_0 \quad \text{and} \quad \sqrt{n}\beta_n \to b = (b_1, b_2, \ldots, b_{d_\beta})' \in [0,\infty]^{d_\beta}. \tag{2.8}$$

Assumption 1 *$\hat\theta_n - \theta_n = o_p(1)$ and $\tilde\theta_n - \theta_n = o_p(1)$ as $n \to \infty$.*

Assumption 2 *(i) $L_n(\theta)$ has continuous left/right partial derivatives of order two on Θ for all $n \geq 1$ (almost surely); (ii) for all deterministic $\epsilon_n \to 0$,*

$$\sup_{\theta\in\Theta:\|\theta-\theta_n\|\leq\epsilon_n} \left\| n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'} - n^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'} \right\| = o_p(1).$$

Assumption 3 *The Hessian satisfies*

$$n^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'} \xrightarrow{p} -\Omega_0 \in \Upsilon.$$

Assumption 4 *The score satisfies*

$$n^{-1/2}\frac{\partial L_n(\theta_n)}{\partial\theta} \xrightarrow{d} N(0,\Sigma_0), \quad \text{with } \Sigma_0 \in \Upsilon$$

Remark 2.1 *Note that Assumptions 1–4 imply that $(\hat\theta_n - \theta_n)$ and $(\tilde\theta_n - \theta_n)$ are $O_p(n^{-1/2})$ (by an extension of Andrews, 1999, Lemma 1 and Theorem 1).*

Running example: Regression. We consider here a setting where the data $\{W_t\}$ are assumed to be i.i.d.; see Remark 3.1 below for the extension to the more general case of time series data. Specifically, we make throughout the following standard assumptions.

Assumption LinIID 1 *For all $\psi \in \Psi$,*

1. *$\{(x_t', \varepsilon_t)'\}_{t=1,2,\ldots}$ are i.i.d., with $\mathbb{E}_\psi[\varepsilon_t|x_t] = 0$ almost surely for all t;*
2. *$\mathbb{E}_\psi[x_t x_t'] \in \Upsilon$ and $\mathbb{E}_\psi[x_t x_t' \varepsilon_t^2] \in \Upsilon$,*
3. *$\mathbb{E}_\psi|x_{j,t}|^{2+\nu} \leq c$ and $\mathbb{E}_\psi|x_{j,t}\varepsilon_t|^{2+\nu} \leq c$ for $j = 1, \ldots, d_\beta$ and constants $c < \infty$ and $\nu > 0$ (not depending on ψ).*

Note that under Assumption 1 and any drifting sequence $\psi_n \in \Psi$ satisfying (2.8), it holds that $\mathbb{E}_{\psi_n}[x_t x_t'] \to \mathbb{E}_{\psi_n}[x_t x_t'] = \Omega_0$ and $\mathbb{E}_{\psi_n}[x_t x_t' \varepsilon_t^2] \to \mathbb{E}_{\psi_0}[x_t x_t' \varepsilon_t^2] = \Sigma_0$. We have the following result that ensures that the high-level Assumptions 1–4 hold for linear regressions.

Proposition 2.1 *Assumption LinIID 1 implies that Assumptions 1–4 hold with* $\Omega_0 = \mathbb{E}_{\psi_0}[x_t x_t']$ *and* $\Sigma_0 = \mathbb{E}_{\psi_0}[x_t x_t' \varepsilon_t^2]$. □

Running example: ARCH. Recall that $\psi = (\theta, \phi)$ where ϕ denotes the joint distribution of F_{t-1} and ε_t. In particular, under the assumptions below, Ω_0 and Σ_0 in Assumptions 3–4 are determined entirely by ψ.

Assumption ARCH 1 *For any* $\psi \in \Psi$ *the process* $\{(y_t^2, F_{t-1}')'\}_{t\in\mathbb{Z}}$ *is stationary and* α*-mixing.*

The assumption is similar to that used in most of the literature on GARCH models, such as Francq and Thieu (2018) where the DGP is assumed to be strictly stationary and ergodic. We here impose the stronger assumption of strong mixing, to apply a weak law of large numbers (LLN) for triangular arrays.

To ensure identification, we make the following assumption.

Assumption ARCH 2 *For any non-zero constant vector* $k \in R^{d_\beta + d_\gamma}$ *and constant* $\tilde{k} \in \mathbb{R}$*, it holds that*

$$\mathbb{P}_\psi(k'(Y_t', X_t')' \neq \tilde{k}) > 0 \quad \text{for all } \psi \in \Psi.$$

Assumption ARCH 3 *For all* $\psi \in \Psi$ *the innovation* ε_t *is independent of the* σ*-field*

$$\mathcal{F}_{t-1} = \sigma\{F_s : s \leq t-1\}. \tag{2.9}$$

Moreover, $\mathbb{E}_\psi[\varepsilon_t^2 - 1] = 0$*, and there exist constants* $\nu_1, c_1 \in (0,\infty)$ *such that* $\mathbb{E}_\psi[|\varepsilon_t|^{4(1+\nu_1)}] \leq c_1$ *for all* $\psi \in \Psi$*. There exists a constant* $\kappa \in (0,\infty)$ *such that* $\mathbb{E}_\psi[\varepsilon_t^4 - 1] = \kappa < \infty$ *for all* $\psi \in \Psi$*.*

Assumption ARCH 4 *There exist constants* $\nu_2, c_2 \in (0,\infty)$ *such that*

$$\mathbb{E}_\psi[\left(y_t^2 \|F_{t-1}\|^3\right)^{1+\nu_2}] \leq c_2 \quad \text{for all } \psi \in \Psi.$$

We note that the space Θ is constrained such that $\{(y_t, F_{t-1}')'\}_{t\in\mathbb{Z}}$ is stationary for any $\theta_0 \in \Theta$. We have the following result.

Proposition 2.2 *Assumptions ARCH 1-4 imply that Assumptions 1-4 hold.* □

2.3 The asymptotic distribution of the LR statistic

The following lemma states the limiting distribution of the LR statistic when some of the nuisance parameters can be on or near the boundary. It extends existing results in the literature (e.g., Theorem 4 of Andrews, 2001) as we consider drifting sequences and hence nuisance parameters local to the boundary of the parameter space. To state the results, define the generic quadratic form

$$Q(\lambda) = \|\lambda - HZ\|^2_{(H\Omega_0^{-1}H')^{-1}}, \tag{2.10}$$

where $\lambda \in \mathbb{R}^{d_\gamma + d_\beta}$ and H is a $(d_\gamma + d_\beta) \times d_\theta$ selection matrix such that $H\theta = (\gamma', \beta')'$. Moreover, Z is $N(0, \Omega_0^{-1}\Sigma_0\Omega_0^{-1})$ distributed, with the matrices Ω_0 and Σ_0 provided by, respectively, by Assumptions 3 and 4.

Lemma 2.1 *Under Assumptions 1–4, and any sequence $\psi_n \in \Psi$ satisfying (2.8),*

$$\mathsf{LR}_n \xrightarrow{d} \mathcal{L}_\infty(b, b)$$

with

$$\mathcal{L}_\infty(x, y) = \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_\beta(x)} Q(\lambda) - \inf_{\lambda \in \Lambda_\gamma \times \Lambda_\beta(y)} Q(\lambda), \tag{2.11}$$

for $x, y \in [0, \infty]^{d_\beta}$, and

$$\Lambda_\gamma = \lim_{n\to\infty} \sqrt{n}(\Theta_\gamma - \gamma_0), \quad \Lambda_\beta(b) = [-b_1, \infty] \times \cdots \times [-b_{d_\beta}, \infty]$$

where Λ_γ is of the form $\Lambda_\gamma = \Lambda_{\gamma,1} \times \cdots \times \Lambda_{\gamma,d_\gamma}$ with $\Lambda_{\gamma,i} = [-\infty, \infty]$ if $\gamma_{0,i}$ is in the interior of $[\gamma_L, \gamma_U]$ and $\Lambda_{\gamma,i} = [0, \infty]$ if $\gamma_{0,i} = \gamma_L$, $i = 1, \ldots, d_\gamma$.

The distribution of the limiting random variable $\mathcal{L}_\infty(b, b)$ approximates the finite-sample distribution of LR_n for $\beta_n = b/\sqrt{n}$. Notice that this distribution is unknown in practice, as it is unknown whether the nuisance parameters in β are interior points or on/near the boundary. Put differently, it is not feasible to consistently estimate the quantiles of $\mathcal{L}_\infty(b, b)$ for use as CVs in this context because b is not consistently estimable along $\{\psi_n\}$ sequences due to the $\sqrt{n}$ scaling of β_n.

Indeed, controlling the asymptotic size of a likelihood ratio test using LR_n as the test statistic requires the use of a CV that asymptotically bounds the rejection probability under all parameter sequences $\psi_n \to \psi_0$ satisfying (2.8) as $n \to \infty$ by the nominal level of the test $\alpha \in (0,1)$ (see, e.g., Andrews and Guggenberger, 2009, or McCloskey, 2017). This motivates us to examine an alternative method of CV construction that can feasibly control the asymptotic size of the test. This we do next.

3 Feasible uniform critical value construction

In this section we detail how to construct uniformly valid critical values for the LR statistic. As we argue below, these CVs are simple to construct, allow to control asymptotic size irrespectively of the nuisance parameters being on the boundary or not, and provide power gains with respect to existing methods.

As outlined in Section 1, and detailed in the next, that as $\mathcal{L}_\infty(b,b)$ is given by (2.11) for $x = y = b$, we suggest to exploit the properties of $\mathcal{L}_\infty(x,y)$ in order to select x and y in a data-driven (and simple) way such that the asymptotic size is bounded above by the user-chosen nominal level α.

We proceed as follows. In Section 3.1 we present a key monotonicity property of $\mathcal{L}_\infty(x,y)$ and discuss initially a naive (inefficient) method to construct valid CVs. In Section 3.2 we introduce some preliminaries needed for constructing our proposed CV. Our main algorithm to construct the CVs is given in Section 3.3, where we also prove the uniform validity of our procedure.

3.1 Monotonicity of $\mathcal{L}_\infty(x,y)$ and some naive CVs

A key propery of $\mathcal{L}_\infty(x,y)$, which we exploit throughout, is given in the following lemma.

Lemma 3.1 *Let* $\underline{b}, b, \bar{b} \in [0,\infty]^{d_\beta}$ *satisfy (element-wise)* $\underline{b} \le b \le \bar{b}$, *then*

$$\mathcal{L}_\infty(b,b) \le \mathcal{L}_\infty(\underline{b},\bar{b}) \ a.s.$$

Given the monotonicity property of Lemma 3.1, a naive and straightforward way of constructing a uniformly valid test is to use a CV based on the distribution of $\mathcal{L}_\infty(0^{d_\beta},\infty^{d_\beta})$, which satisfies $\mathcal{L}_\infty(b,b) \leq \mathcal{L}_\infty(0^{d_\beta}, \infty^{d_\beta})$. However, such a choice of CVs leads to a very conservative test. Alternatively, one may use the $1-\alpha$ quantile of $\mathcal{L}_\infty(b^*, b^*)$, where b^* is the maximizer across $[0,\infty]^{d_\beta}$ of all $1-\alpha$ quantiles of $\mathcal{L}_\infty(b,b)$, in Andrews and Guggenberger (2009). However, also this choice of CV can lead to a conservative test with lower power. Perhaps more importantly, b^* is computationally prohibitive to compute when the number of nuisance parameters $d_\beta > 2$. To reduce the conservative nature of these CVs, McCloskey (2017) suggests to use the $1-\alpha+\eta$ quantile of $\mathcal{L}_\infty(\tilde{b}, \tilde{b})$ for some $\eta \in (0,\alpha)$, where $\tilde{b}$ is the maximizer across a $(1-\eta)$-level confidence set for b. Unfortunately, this proposal suffers a similar computational drawback to that of Andrews and Guggenberger (2009) when $d_\beta > 2$.

We finally note that the shrinkage-based bootstrap of Cavaliere, Nielsen, Pedersen and Rahbek (2022) essentially seeks to choose between $\mathcal{L}_\infty(0,0)$ or $\mathcal{L}_\infty(\infty,\infty)$ (component-wise) in a data-driven way. However, it fails to control size uniformly since there may exist values $b \in (0,\infty)$ such that $\mathcal{L}_\infty(b,b) \geq \mathcal{L}_\infty(0,0)$.

3.2 Prerequisites and additional assumptions

In order to construct computationally-feasible and uniformly-valid CVs, we require two main ingredients, which are very simple to satisfy in applications. First, we need a consistent estimators of the covariance matrices Ω_0 and Σ_0, see Assumptions 3–4. Second, we need the construction of a consistent and asymptotically Gaussian estimator $\check{\beta}_n$ of β_n with asymptotic covariance matrix given by $\Sigma_\beta = H_\beta \Omega_0^{-1} \Sigma_0 \Omega_0^{-1} H_\beta'$ where Ω_0 is defined in Assumption 3 and H_β is the selection matrix satisfying $H_\beta \theta = \beta$. Moreover, we require a consistent estimator $\check{\Sigma}_{\beta,n}$ of the covariance matrix Σ_β. We formalize these requirements through the following assumptions, which hold for any sequence $\{\psi_n\}$ in Ψ satisfying (2.8).

Assumption 5 *There exists a matrix* $\hat{\Sigma}_n$ *such that* $\hat{\Sigma}_n \xrightarrow{p} \Sigma_0$.

Assumption 6 *There exists estimators* $\check{\beta}_n$ *and* $\check{\Sigma}_{\beta,n}$ *which satisfy (i)* $\sqrt{n}(\check{\beta}_n - \beta_n) \xrightarrow{d} N(0, \Sigma_\beta)$, *with* $\Sigma_\beta = H_\beta \Omega_0^{-1} \Sigma_0 \Omega_0^{-1} H_\beta'$, *and (ii)* $\check{\Sigma}_{\beta,n} \xrightarrow{p} \Sigma_\beta$.

As a simple example of an estimator satisfying Assumption 6 in general, consider the one-step iterated Newton-Raphson estimator for θ:

$$\check{\theta}_n = \hat{\theta}_n - \left(\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1} \frac{\partial L_n(\hat{\theta}_n)}{\partial\theta}. \tag{3.1}$$

As observed by Ketz (2018), by definition, $\check{\theta}_n$ may not belong to the parameter space Θ, and hence $L_n(\check{\theta}_n)$ may not even be well-defined. We have the following lemma due to Ketz (2018).

Lemma 3.2 *Under any sequence $\psi_n \in \Psi$ satisfying (2.8) and Assumptions 1–4, with $\check{\theta}_n$ given by (3.1),*

$$\sqrt{n}(\check{\theta}_n - \theta_n) \xrightarrow{d} N(0, \Omega_0^{-1}\Sigma_0\Omega_0^{-1}).$$

In particular, we have that $\check{\beta}_n = H_\beta\check{\theta}_n$ is asymptotically normal,

$$\sqrt{n}(\check{\beta}_n - \beta_n) \xrightarrow{d} N(0, \Sigma_\beta), \tag{3.2}$$

where $\Sigma_\beta = H_\beta\Omega_0^{-1}\Sigma_0\Omega_0^{-1}H_\beta'$. If, in addition, Assumption 5 holds, then

$$\check{\Sigma}_{\beta,n} = H_\beta \left(n^{-1}\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1} \hat{\Sigma}_n \left(n^{-1}\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1} H_\beta' \xrightarrow{p} \Sigma_\beta. \tag{3.3}$$

Hence, for the one-step iterated Newton-Raphson estimator $\check{\theta}_n$ of (3.1) it holds that Assumptions 1–5 imply Assumption 6. We illustrate this in terms of the two running examples.

Running example: Regression. To verify Assumptions 5 and 6 for the linear regression example, we make the following additional assumptions.

Assumption LinIID 2 *For all $\psi \in \Psi$ and some constants $\nu, c > 0$, $\mathbb{E}_\psi[|x_{i,t}, x_{j,t}, x_{k,t}x_{l,t}|^{1+\nu}] \leq c$ for i,$j, k, l = 1, \ldots, d_\theta$.*

The following Proposition states that Assumptions 5–6 hold for the linear regression example when choosing $\hat{\Sigma}_n$ as the Eicker-White heteroskedasticity-robust estimator[1],

$$\hat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t^2 x_t x_t', \quad \hat{\varepsilon}_t = y_t - x_t'\hat{\theta}_{LS}. \tag{3.4}$$

[1]Note that one could alternatively define the estimator in terms of the the constrained residuals, that is, $\hat{\varepsilon}_t = y_t - x_t'\hat{\theta}_n$.

Proposition 3.1 *Suppose* $\check{\beta}_n = \hat{\beta}_{LS}$ *and* $\check{\Sigma}_{\beta,n}$ *is equal to the upper* $d_\beta \times d_\beta$ *block of* $S_{xx}^{-1}\hat{\Sigma}_n S_{xx}^{-1}$ *with* $\hat{\Sigma}_n$ *given by (3.4). Then Assumptions LinIID 1–2 imply Assumptions 5–6.*

Remark 3.1 *The linear regression example can easily be extended to time series data. In particular, under appropriate strong mixing conditions on* $\{W_t\}_{t\in\mathbb{Z}}$ *and replacing* $\hat{\Sigma}_n$ *with a HAC estimator (Newey and West, 1987), it is possible to verify Assumptions 1–5.* □

Running example: ARCH. Under Assumptions ARCH 1-4 it holds that $\Sigma_0 = (\kappa/2)\Omega_0$; hence, we can estimate Σ_0 using $\hat{\Sigma}_n = (\hat{\kappa}_n/2)\hat{\Omega}_n$ with

$$\hat{\Omega}_n = -n^{-1}\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}, \qquad \hat{\kappa}_n = n^{-1}\sum_{t=1}^{n}(\hat{\varepsilon}_t^4 - 1)$$

where $\hat{\varepsilon}_t = y_t/\hat{\sigma}_t(\hat{\theta}_n)$. Moreover, let $\check{\beta}_n = H_\beta\check{\theta}_n$ with $\check{\theta}_n$ given by (3.1), and finally let

$$\check{\Sigma}_{\beta,n} = (\hat{\kappa}_n/2)H_\beta\hat{\Omega}_n^{-1}H_\beta'. \tag{3.5}$$

Proposition 3.2 *With* $\check{\beta}_n$ *and* $\check{\Sigma}_{\beta,n}$ *as defined above, Assumptions ARCH 1-4 imply Assumptions 5–6.* □

3.3 Uniformly valid CV

We can now finally state our main algorithm to construct uniformly valid CVs, which are denoted as $CV_{\alpha,n}$, where $\alpha \in (0,1)$ denotes the nominal significance level. As anticipated, it relies on the asymptotically normal estimator $\check{\beta}_n$ and the consistent covariance matrix estimator $\check{\Sigma}_{\beta,n}$ in Assumption 6.

Algorithm 1

With $\hat{\theta}_n$, $\hat{\Omega}_n = -n^{-1}\partial^2 L_n(\hat{\theta}_n)/\partial\theta\partial\theta'$, $\hat{\Sigma}_n$, $\check{\beta}_n$, $\check{\Sigma}_{\beta,n}$ *and the nominal significance level* $\alpha \in (0,1)$ *as inputs:*

1. *Choose some* $\eta \in (0,\alpha)$;
2. *Compute* $\check{\Omega}_{\beta,n} = \operatorname{diag}(\check{\Sigma}_{\beta,n})^{-1/2}\check{\Sigma}_{\beta,n}\operatorname{diag}(\check{\Sigma}_{\beta,n})^{-1/2}$;
3. *Compute* $\check{q}_{1-\eta,n}$ *as the* $(1-\eta)$*-quantile of* $\max_{i=1,\dots,d_\beta}|\check{Z}_{\beta,i}|$ *with* $\check{Z}_\beta \overset{d}{=} N(0,\check{\Omega}_{\beta,n})$;

4. *Compute* $b_{L,n} = (b_{L,n,1}, \ldots, b_{L,n,d_\beta})'$ *and* $b_{U,n} = (b_{U,n,1}, \ldots, b_{U,n,d_\beta})'$ *by setting* $b_{L,n,i} = \max\{0, \check{b}_{L,n,i}\}$ *and* $b_{U,n,i} = \max\{0, \check{b}_{U,n,i}\}$ *for* $i = 1, \ldots, d_\beta$*, where*

$$\check{b}_{L,n} = \sqrt{n}\check{\beta}_n - \check{q}_{1-\eta,n}\,\mathrm{diagv}(\check{\Sigma}_{\beta,n})^{1/2},\ \check{b}_{U,n} = \sqrt{n}\check{\beta}_n + \check{q}_{1-\eta,n}\,\mathrm{diagv}(\check{\Sigma}_{\beta,n})^{1/2};$$

5. *Compute* $\mathsf{CV}_{\alpha,n}$ *as the* $1 - \alpha + \eta$ *quantile of* $\mathcal{L}_{\infty,n}(b_{L,n}, b_{U,n})$*, where* $\mathcal{L}_{\infty,n}(\cdot,\cdot)$ *is defined as* $\mathcal{L}_\infty(\cdot,\cdot)$ *in (2.11) with* Ω_0 *and* Σ_0 *replaced by* $\check{\Omega}_n$ *and* $\hat{\Sigma}_n$*, respectively.*

Notice that the quantiles in Steps 3 and 5 of Algorithm 1 can be computed straightforwardly by simulation. Notice also that the elements of $b_{L,n}$ and $b_{U,n}$ are maximized against zero in order to ensure that $b_{L,n}$ and $b_{U,n}$ obeys the lower bounds of b.

We can now state our main theorem, where we establish the uniform asymptotic size control of the LR test based on our proposed critical value, $CV_{\alpha,n}$.

Theorem 3.3 *Let* $\mathsf{CV}_{\alpha,n}$ *be constructed as in Algorithm 1. Then under Assumptions 1–6,*

$$\limsup_{n\to\infty} \sup_{\psi\in\Psi} \mathbb{P}_\psi\left(\mathsf{LR}_n \geq \mathsf{CV}_{\alpha,n}\right) \leq \alpha.$$

We emphasize that the underlying assumptions of Theorem 3.3 are intuitive and typically possible to very for a given statistical model. In particular, have that the critical values determined by Algorithm 1 are valid in terms of the linear regression and the ARCH models due to Propositions 3.1 and 3.2, respectively. The result states that Algorithm 1 provides asymptotically valid CVs for any choice of $\eta \in (0, \alpha)$. We note that $1 - \eta$ (approximately) quantifies the probability of the event $b \in [b_{L,n}, b_{U,n}]$. The parameter η is user-chosen and following the existing body of literature (e.g., McCloskey, 2017) we recommend choosing $\eta = \alpha/10$.

4 Simulation experiments

In this section, we consider the performance of our proposed LR test. Section 4.1 considers the linear regression with positivity-constrained coefficients as in the running example, whereas Section 4.2 considers the ARCH model.

4.1 Linear regression with positivity constraints

Consider the linear regression model, given by

$$y_t = \gamma x_{t,1} + \beta x_{t,2} + \varepsilon_t, \quad t = 1, \ldots, n,$$

where for $x_t = (x_{t,1}, x_{t,2})'$, $\{(x_t', \varepsilon_t)'\}_{t=1,2,\ldots}$ an i.i.d. process with

$$\begin{pmatrix} x_t \\ \varepsilon_t \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Omega & 0 \\ 0 & 1 \end{pmatrix}\right),$$

and $\theta = (\gamma, \beta')'$ with $\gamma, \beta \in [0, \infty)$. The matrix Ω is a correlation matrix given by

$$\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

We seek to test the hypothesis $\mathsf{H}_0 : \gamma = 0$ against $\gamma > 0$.

For comparison, we report the rejection frequencies for the LR test where one ignores that β is constrained. Here the CV is given by 2.71, the 90th percentile of the χ_1^2 distribution, which is motivated by (erroneously) approximating the distribution of the LR statistic by its asymptotic distribution at the boundary point $\beta = 0$, $(\max\{0, N(0,1)\})^2$ (labelled "LR" in the tables). We also compare with the conditional LR ("CLR" in the tables) test proposed by Ketz (2018). All tests are carried out at a nominal level of $\alpha = 5\%$, and for the uniform CV construction $\eta = \alpha/10$. For the uniform CV construction as well as the CLR test, the CVs are determined by means of simulation, making use of 10,000 draws. All rejection frequencies are based on 10,000 Monte Carlo replications.

Tables 1-4 report the rejection frequencies of our proposed test for different values of γ, β, ρ, and n. Specifically, we consider the cases $\gamma, \beta \in \{0, 0.1\}$, $\rho \in \{-0.95, -0.75, 0.5, 0, 0.5, 0.75, 0.95\}$ and sample sizes $n \in \{100, 250, 500, 1000\}$.

Tables 1 and 2 contain the rejection frequencies under H_0.

[Tables 1 and 2 around here]

Our test appears to control size. It tends to be conservative whenever $(\max(N(0,1), 0)^2$ yields conservative CVs, that is for correlations $\rho \geq 0.5$. Importantly, it controls size when $(\max(N(0,1),0)^2$ does not, that is for correlations $\rho \leq -0.5$. The CLR test has rejection frequencies around α across all correlations.

Tables 3 and 4 contains the rejection frequencies under the alternative.

[Tables 3 and 4 around here]

Our proposed method appears to have attractive rejection frequencies under most alternatives, and in particular it has higher rejection frequencies than the CLR approach for correlations $\rho \leq -0.75$.

4.2 ARCH

In this section we consider the ARCH model and seek to evaluate the performance of our proposed method when testing for the presence of an explanatory covariate when another covariate may be present. Specifically, consider the model

$$y_t = \sigma_t \varepsilon_t, \quad t = 1, \ldots, n,$$
$$\sigma_t^2 = \delta_1 + \delta_2 y_{t-1}^2 + \gamma x_{t-1,1} + \beta x_{t-1,2},$$

where $\delta_1 > 0$, $\delta_2, \gamma, \beta \geq 0$, and $\{\varepsilon_t\}_{t=1,\ldots,n}$ is an i.i.d. process with $\varepsilon_t \sim N(0,1)$. We consider testing $\mathsf{H}_0 : \gamma = 0$, against $\gamma > 0$.

In terms of the covariates, we follow the simulation design in Nielsen, Pedersen, Rahbek and Thorsen (2024) and let $(x_{t,1}, x_{t,2}) = (\exp(v_{t,1}), \exp(v_{t,2}))$ with $V_t = (v_{1,t}, v_{2,t})'$ a bivariate autoregression with correlated innovations satisfying $V_t = aV_{t-1} + \epsilon_t$, $t = 1, \ldots, n$. Here $\{\epsilon_t\}_{t=1}^n$ is an i.i.d. $N_2(0, \Sigma)$ process with

$$\Sigma = b\left(1 - a^2\right) \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix}$$

independent of $\{z_t\}_{t=1}^n$, $\rho_{12} \in (-1, 1)$, $a = 0.9$ and $b = 0.5$. The simulated realizations of the processes for y_t and $(x_{t,1}, x_{t,2})$ make use of a burn-in period of 1000 observations.

For the experiment we assume that it is known to the researcher that the true value of the ARCH coefficient δ_2 is not near its boundary of zero, so that the only parameter that potentially causes a discontinuity in the null distribution is β. We report the rejection frequencies for $n = 1000$ observations, parameter values $\gamma, \beta \in \{0, 0.01, 0.05, 0.1, 0.25\}$, $\rho_{12} \in \{-0.95, -0.75, 0.5, 0, 0.5, 0.75, 0.95\}$, and compare with the standard LR with CVs derived at the boundary as well as the CLR test as in the previous section.

Table 5 contains rejection frequencies under H_0 for different values of β and ρ_{12}. We note that similar to the findings for the linear regression case in the previous section, our proposed method has rejection frequencies less than α for all combinations of parameter values. Similar to the linear regression case, the standard LR test overrejects for the case of $\beta = 0$ and small values of ρ_{12}.

[Tables 5 and 6 around here]

Table 6 contains the rejection frequencies for the alternatives $\gamma \in \{0.01, 0.05, 0.1, 0.25\}$ with $\beta = 0$. Our proposed method performs well in terms of rejecting H_0, and performs comparably to the CLR test for most combinations of the parameter values.

Summarizing the findings in Sections 4.1 and 4.2, we have that our proposed method yields attractive rejection frequencies in most settings and yield more powerful tests than the CLR approach in cases with extreme correlations between covariates.

5 Empirical illustration

In this section we consider an empirical illustration of our proposed test. We consider ARCH models for daily returns of various stock indices and test for the presence of ARCH effects and spillovers from the U.S. stock market. Inspired by the HAR model of Corsi (2009), and similar to Nielsen et al. (2024, Section 5), we include lagged Realized Volatility (RV) covariates to account for potential high persistence in the conditional variance of the index returns. With the *daily* index return given by y_t in (2.4), let

$$\sigma_t^2 = \delta + \beta_{ARCH} y_{t-1}^2 + \beta_{RV} RV_{t-1} + \beta_W RV_{W,t-1} + \beta_M RV_{M,t-1} + \beta_{SPX} SPX_{t-1}^2,$$

with $\delta > 0$ and $\beta_{ARCH}, \beta_{RV}, \beta_M, \beta_{SPX} \geq 0$. The variable RV_t is the RV of the index based on 5-minutes intraday returns at day t, $RV_{W,t} = 5^{-1} \sum_{i=0}^{4} RV_{t-i}$ is the weekly average RV, and $RV_{W,t} = 22^{-1} \sum_{i=0}^{21} RV_{t-i}$ is the monthly average RV. Lastly, the variable SPX_t is the (continuously compounded, close-to-close) return on the S&P 500 index at day t.

We refer to Nielsen et al. (2024, Section 5) for a discussion and motivation for this type of model. Based on data retrieved from the Oxford Man Realized Library covering the period 3 January 2000 to 27 June 2018, we analyze the following indices: Australian ASX All Ordinaries (AORD), Belgian BEL 20 (BFX), Spanish IBEX 35 (IBEX), IPC Mexico (MXX), Indian NIFTY50 (NSEI) and Danish OMX C20 (OMX) (notice that OXM C20 begins on 3 October 2005). For each of these indices, we test the hypothesis for no ARCH, $\mathsf{H}_{ARCH} : \beta_{ARCH} = 0$, and the hypothesis of no spillovers from the U.S. stock market, $\mathsf{H}_{SPX} : \beta_{SPX} = 0$. The tests are carried out under the assumption that the true value δ_0 is away from its lower bound, $\delta_0 > \delta_L$, but no assumptions are imposed on the remaining nuisance parameters.

Table 7 contains point estimates of the model parameters for each index series. Moreover, it contains the values of the LR and CLR statistics along with CVs based on 5% nominal levels. Based on the LR test, for all of the series except for the Indian NSEI index, we cannot reject H_{ARCH}. On the contrary we reject H_{SPX} for all series except for the Danish OMX. Note that if one uses as CV, 2.71, based on the $\max\{N(0,1),0\}^2$-distribution, one would reject H_{SPX} also for the OMX series.

[Table 7 around here]

In short, based on our proposed critical values, we find evidence for no ARCH effects in most of the index return series, whereas most of the series appear subject to spillovers from the U.S. stock market.

6 Conclusions

This paper proposes a novel and computationally efficient method for constructing CVs for LR tests that achieve uniform asymptotic size control, even when nuisance parameters lie on or near the boundary of their parameter space. The key innovation lies in using confidence bounds for an asymptotically Gaussian approximation of a nuisance parameter estimator and exploiting the monotonicity properties of the two components of the LR statistic's asymptotic distribution. This allows for the formation of valid CVs via straightforward Monte Carlo simulation, without the need for intensive optimization or conservative least-favorable approaches. The method remains tractable in settings for which the nuisance parameter is not low-dimensional and is broadly applicable, including to models such as constrained regressions and ARCH specifications, offering the first uniformly valid test for the latter.

References

Andrews, D.W.K. (2001): "Testing when a parameter is on the boundary of the maintained hypothesis", *Econometrica*, vol. 69, 683–734.

Andrews, D.W.K. and Cheng, X. (2012),"Estimation and inference with weak, semi-strong, and strong identification", *Econometrica*, vol. 80, 2153–2211.

Andrews, D.W.K. and Guggenberger, P. (2009): "Hybrid and size-corrected subsampling methods", *Econometrica*, vol. 77, 721–762.

Berger, R. L. and Boos, D. D. (1994): "P values maximized over a confidence set for the nuisance parameter", *Journal of the American Statistical Association*, vol. 89, 1012–1016.

Berry, S., Levinsohn, J., and Pakes, A. (1995): "Automobile prices in market equilibrium", *Econometrica*, vol. 63, 841–890.

Cavaliere, G., Nielsen, H. B., Pedersen, R. S., and Rahbek, A. (2022): "Bootstrap inference on the boundary of the parameter space, with application to conditional volatility models", *Journal of Econometrics*, vol. 227, 241–263.

Cavaliere, G., Nielsen, H. B., and Rahbek, A. (2017): "On the consistency of bootstrap testing for a parameter on the boundary of the parameter space", *Journal of Time Series Analysis*, vol. 38, 513–534.

Cavaliere, G., Perera, I., and Rahbek, A. (2024): "Specification tests for GARCH processes", *Journal of Business and Economic Statistics*, vol. 42, 197–214.

Corsi, F. (2009): "A simple approximate long-memory model of realized volatility", *Journal of Financial Econometrics*, vol. 7, 174–196.

Drton, M. (2009): "Likelihood ratio tests and singularities", *Annals of Statistics*, vol. 37, 979–1012.

Fan, Y. and Shi, X. (2023): "Wald, QLR, and score tests when the parameters are subject to linear inequality constraints", *Journal of Econometrics*, vol. 235, 2005-2026.

Francq, C. and Thieu, L. Q. (2019): "QML inference for volatility models with covariates", *Econometric Theory*, vol. 35, 37–72.

Francq, C., and Zakoïan, J.-M. (2009), "Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons", *Journal of the American Statistical Association*, vol. 104, 313–324.

Geyer, C. J. (1994),"On the asymptotics of constrained M-estimators", *The Annals of Statistics*, vol. 22, 1993–2010.

Hong, H. and Li, J. (2020),"The numerical bootstrap", *The Annals of Statistics*, vol. 48, 397–412.

Ketz, P. (2018): "Subvector inference when the true parameter vector may be near or at the boundary", *Journal of Econometrics*, vol. 207, 285–306.

Ketz, P. and McCloskey, A. (2023): "Short and simple confidence intervals when the directions of some effects are known", *Review of Economics and Statistics*, forthcoming.

Li, J. (2025),"The proximal bootstrap for constrained estimators", *Journal of Statistical Planning and Inference*, vol. 236, 106245.

McCloskey, A. (2017): "Bonferroni-based size-correction for nonstandard testing problems", *Journal of Econometrics*, vol. 200, 17–35.

Mitchell, J. D., Allman, E. S., and Rhodes, J. A. (2019): "Hypothesis testing near singularities and boundaries", *Electronic Journal of Statistics*, vol. 13, 2150–2193.

Newey, W. K. and McFadden, D. L. (1994): "Large sample estimation and hypothesis testing". In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, vol. 4, 2111–2245.

Newey, W. K. and West, K. D. (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", *Econometrica*, vol. 55, 703–708.

Nielsen, H. B., Pedersen, R. S., Rahbek, A., and Thorsen, S. N.,"Testing in GARCH-X models: Boundary, correlations and bootstrap theory", *Journal of Time Series Analysis*, forthcoming.

Self, S. G. and Liang, K.-Y. (1987),"Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions", *Journal of the American Statistical Association*, vol. 82, 605–610.

Shapiro, A. (1989),"Asymptotic properties of statistical estimators in stochastic programming", *The Annals of Statistics*, vol. 17, 841–858.

Silvapulle, M. J. (1996),"A test in the presence of nuisance parameters", *Journal of the American Statistical Association*, vol. 91, 1690–1693.

Silvapulle, M. J. and Silvapulle, P. (1995),"A score test against one-sided alternatives", *Journal of the American Statistical Association*, vol. 90, 342–349.

Table 1: Rejection Frequencies under null hypothesis, $(\gamma, \beta) = (0, 0)$

n	LR	CLR	LR-uniform
	$\rho = -0.95$		
100	0.0998	0.0520	0.0414
250	0.1050	0.0526	0.0397
500	0.1054	0.0536	0.0422
1000	0.1040	0.0502	0.0424
	$\rho = -0.75$		
100	0.0813	0.0527	0.0352
250	0.0873	0.0545	0.0359
500	0.0820	0.0513	0.0332
1000	0.0867	0.0529	0.0367
	$\rho = -0.5$		
100	0.0727	0.0523	0.0299
250	0.0684	0.0505	0.0297
500	0.0753	0.0528	0.0309
1000	0.0677	0.0482	0.0250
	$\rho = 0$		
100	0.0520	0.0545	0.0223
250	0.0489	0.0512	0.0183
500	0.0510	0.0513	0.0191
1000	0.0482	0.0498	0.0190
	$\rho = 0.5$		
100	0.0293	0.0547	0.0114
250	0.0283	0.0501	0.0105
500	0.0288	0.0512	0.0099
1000	0.0286	0.0519	0.0103
	$\rho = 0.75$		
100	0.0142	0.0552	0.0049
250	0.0143	0.0584	0.0041
500	0.0138	0.0550	0.0043
1000	0.0152	0.0577	0.0046
	$\rho = 0.95$		
100	0.0014	0.0602	0.0011
250	0.0014	0.0577	0.0010
500	0.0015	0.0634	0.0011
1000	0.0013	0.0603	0.0010

Table 2: Rejection Frequencies under null hypothesis, $(\gamma, \beta) = (0, 0.1)$

n	LR	CLR	LR-uniform
	$\rho = -0.95$		
100	0.0562	0.0537	0.0224
250	0.0540	0.0536	0.0205
500	0.0495	0.0493	0.0217
1000	0.0488	0.0507	0.0237
	$\rho = -0.75$		
100	0.0560	0.0553	0.0226
250	0.0527	0.0530	0.0235
500	0.0547	0.0554	0.0294
1000	0.0510	0.0507	0.0369
	$\rho = -0.5$		
100	0.0554	0.0544	0.0226
250	0.0514	0.0519	0.0250
500	0.0470	0.0468	0.0257
1000	0.0488	0.0487	0.0339
	$\rho = 0$		
100	0.0522	0.0543	0.0213
250	0.0505	0.0512	0.0235
500	0.0497	0.0514	0.0247
1000	0.0526	0.0523	0.0302
	$\rho = 0.5$		
100	0.0423	0.0538	0.0155
250	0.0499	0.0544	0.0195
500	0.0493	0.0504	0.0170
1000	0.0460	0.0463	0.0169
	$\rho = 0.75$		
100	0.0288	0.0543	0.0083
250	0.0367	0.0487	0.0113
500	0.0445	0.0502	0.0148
1000	0.0496	0.0510	0.0185
	$\rho = 0.95$		
100	0.0048	0.0526	0.0017
250	0.0108	0.0536	0.0028
500	0.0226	0.0547	0.0065
1000	0.0364	0.0543	0.0085

Table 3: Rejection Frequencies under alternative hypothesis, $(\gamma, \beta) = (0.1, 0)$

n	LR	CLR	LR-uniform
	$\rho = -0.95$		
100	0.3214	0.0947	0.1811
250	0.5331	0.1250	0.3518
500	0.7741	0.1780	0.6037
1000	0.9504	0.2607	0.8811
	$\rho = -0.75$		
100	0.3232	0.1707	0.1870
250	0.5326	0.2726	0.3529
500	0.7638	0.4411	0.6022
1000	0.9432	0.6702	0.8702
	$\rho = -0.5$		
100	0.2947	0.2150	0.1676
250	0.5257	0.3929	0.3487
500	0.7499	0.6147	0.5870
1000	0.9443	0.8611	0.8758
	$\rho = 0$		
100	0.2665	0.2708	0.1508
250	0.4712	0.4774	0.3061
500	0.7124	0.7146	0.5519
1000	0.9302	0.9305	0.8542
	$\rho = 0.5$		
100	0.1720	0.2246	0.0902
250	0.3474	0.3961	0.1997
500	0.5695	0.6088	0.3966
1000	0.8498	0.8599	0.7174
	$\rho = 0.75$		
100	0.0958	0.1660	0.0432
250	0.2108	0.2783	0.1028
500	0.3851	0.4332	0.2238
1000	0.6554	0.6733	0.4730
	$\rho = 0.95$		
100	0.0106	0.0921	0.0036
250	0.0290	0.1232	0.0105
500	0.0752	0.1737	0.0237
1000	0.1796	0.2560	0.0620

Table 4: Rejection Frequencies under alternative hypothesis $(\gamma, \beta) = (0.1, 0.1)$

n	LR	CLR	LR-uniform
	$\rho = -0.95$		
100	0.1474	0.0927	0.0691
250	0.1804	0.1229	0.0803
500	0.2267	0.1716	0.1114
1000	0.3168	0.2598	0.1666
	$\rho = -0.75$		
100	0.1991	0.1690	0.1010
250	0.3152	0.2853	0.1820
500	0.4570	0.4335	0.2931
1000	0.6855	0.6729	0.5117
	$\rho = -0.5$		
100	0.2294	0.2190	0.1251
250	0.4054	0.3947	0.2557
500	0.6235	0.6196	0.4553
1000	0.8668	0.8649	0.7586
	$\rho = 0$		
100	0.2588	0.2659	0.1466
250	0.4656	0.4719	0.3103
500	0.7107	0.7118	0.5699
1000	0.9313	0.9314	0.8887
	$\rho = 0.5$		
100	0.2071	0.2269	0.1044
250	0.3861	0.3921	0.2311
500	0.6121	0.6135	0.4495
1000	0.8640	0.8639	0.7921
	$\rho = 0.75$		
100	0.1416	0.1700	0.0630
250	0.2724	0.2818	0.1427
500	0.4353	0.4369	0.2721
1000	0.6657	0.6660	0.5083
	$\rho = 0.95$		
100	0.0305	0.0964	0.0099
250	0.0822	0.1269	0.0255
500	0.1567	0.1712	0.0623
1000	0.2527	0.2527	0.1272

Table 5: Rejection Frequencies under null hypothesis

β_2	LR	CLR	LR-Uniform
	$\rho_{12} = -0.95$		
0	0.07814	0.03445	0.03094
0.01	0.05682	0.04196	0.0257
0.05	0.0504	0.03612	0.04296
0.1	0.05299	0.03883	0.05491
0.25	0.04597	0.04638	0.05346
	$\rho_{12} = -0.75$		
0	0.07296	0.03733	0.03252
0.01	0.04922	0.0389	0.02226
0.05	0.05032	0.03988	0.04359
0.1	0.05171	0.04084	0.05311
0.25	0.05149	0.04218	0.05462
	$\rho_{12} = -0.5$		
0	0.06602	0.03927	0.02835
0.01	0.05276	0.04005	0.02233
0.05	0.05114	0.03901	0.04231
0.1	0.04819	0.03945	0.0503
0.25	0.04596	0.03941	0.0504
	$\rho_{12} = 0$		
0	0.04871	0.03818	0.01854
0.01	0.04683	0.03913	0.02071
0.05	0.05045	0.03964	0.03844
0.1	0.05148	0.04187	0.05158
0.25	0.04842	0.03589	0.04882
	$\rho_{12} = 0.5$		
0	0.02877	0.05214	0.01233
0.01	0.04805	0.04324	0.01992
0.05	0.04851	0.04021	0.0245
0.1	0.04613	0.03792	0.03592
0.25	0.04931	0.03981	0.04771
	$\rho_{12} = 0.75$		
0	0.01475	0.0620	0.00612
0.01	0.03772	0.04913	0.01501
0.05	0.0488	0.0434	0.0189
0.1	0.0504	0.0454	0.0230
0.25	0.0482	0.0430	0.0385
	$\rho_{12} = 0.95$		
0	0.0009022	0.05704	0.001403
0.01	0.0106	0.05932	0.003701
0.05	0.0421	0.0489	0.0118
0.1	0.0538	0.0537	0.0217
0.25	0.0481	0.0481	0.0188

Table 6: Rejection Frequencies for under various alternatives

γ	LR	CLR	LR-uniform
	$\rho_{12} = -0.95$		
0	0.08134	0.0379	0.03298
0.01	0.4380	0.2167	0.2813
0.05	0.9902	0.7901	0.9649
0.1	0.9998	0.9114	0.9965
0.25	1.0000	0.9568	0.9981
	$\rho_{12} = -0.75$		
0	0.0701	0.03691	0.02888
0.01	0.4377	0.2610	0.2872
0.05	0.9891	0.8976	0.9687
0.1	1.0000	0.9714	0.9982
0.25	1.0000	0.9856	0.9986
	$\rho_{12} = -0.5$		
0	0.06329	0.03815	0.02544
0.01	0.4230	0.2956	0.2765
0.05	0.9889	0.9527	0.9688
0.1	0.9999	0.9939	0.9988
0.25	1.0000	0.9961	0.9990
	$\rho_{12} = 0$		
0	0.0478	0.03818	0.01914
0.01	0.3816	0.3355	0.2459
0.05	0.9875	0.9819	0.9670
0.1	0.9999	0.9997	0.9990
0.25	1.0000	0.9999	0.9994
	$\rho_{12} = 0.5$		
0	0.03165	0.05479	0.01202
0.01	0.2879	0.3144	0.1702
0.05	0.9655	0.9578	0.9191
0.1	0.9988	0.9985	0.9967
0.25	1.0000	1.0000	1.0000
	$\rho_{12} = 0.75$		
0	0.01686	0.0588	0.006924
0.01	0.1721	0.2313	0.08815
0.05	0.8592	0.8464	0.7527
0.1	0.9821	0.9784	0.9581
0.25	0.9996	0.9994	0.9990
	$\rho_{12} = 0.95$		
0	0.002107	0.06493	0.001706
0.01	0.02321	0.1231	0.01021
0.05	0.3389	0.3702	0.1745
0.1	0.5843	0.5847	0.4011
0.25	0.8056	0.8030	0.6670

Table 7: Results for the empirical illustration

	AORD	BFX	IBEX	MXX	NSEI	OMX
δ	0.03	0.05	0.08	0.25	0.35	0.27
β_{ARCH}	0.01	0.01	0.00	0.08	0.16	0.00
β_{RV}	0.09	0.63	0.62	0.30	0.39	0.49
β_W	0.69	0.79	0.52	0.37	0.38	0.58
β_M	0.34	0.00	0.17	0.72	0.22	0.05
β_{SPX}	0.19	0.07	0.09	0.07	0.15	0.06
LR(H_{ARCH})	0.66	0.66	0.00	16.78	106.42	0.00
Uniform CV	6.58	6.66	8.39	16.98	20.44	18.50
LR(H_{SPX})	441.02	23.24	21.79	22.07	59.99	6.05
Uniform CV	7.57	7.29	8.52	16.95	20.21	18.76
Observations	4757	4802	4772	4729	4670	3259

Supplemental Appendix

A Proofs

A.1 Proofs of main results

Proof of Lemma 2.1

We first introduce some notation. Let $b^{(1)} = (b_1^{(1)}, \ldots, b_{d_\beta}^{(1)})'$ and $b^{(2)} = (b_1^{(2)}, \ldots, b_{d_\beta}^{(2)})'$ satisfy

$$b_i^{(1)} = \mathbb{I}(b_i < \infty) b_i, \ b_i^{(2)} = \mathbb{I}(b_i = \infty) b_i$$

for $i = 1, \ldots, d_\beta$, and where we work under the convention that $0 \times \infty = 0$. Note that $b = b^{(1)} + b^{(2)}$. Likewise, let $\beta_n^{(1)} = (\beta_{n,1}^{(1)}, \ldots, \beta_{n,d_\beta}^{(1)})'$ and $\beta_n^{(2)} = (\beta_{n,1}^{(2)}, \ldots, \beta_{n,d_\beta}^{(2)})'$ satisfy

$$\beta_{n,i}^{(1)} = \mathbb{I}(b_i < \infty)\beta_{n,i}, \quad \beta_{n,i}^{(2)} = \mathbb{I}(b_i = \infty)\beta_{n,i}$$

for $i = 1, \ldots, d_\beta$. Define

$$\theta_n^{(1)} = (0'_{d_\gamma}, \beta_n^{(1)\prime}, 0'_{d_\delta})' \quad \text{and} \quad \theta_n^{(2)} = (\gamma_0', \beta_n^{(2)\prime}, \delta_0')'. \tag{A.1}$$

In particular, $\theta_n = \theta_n^{(1)} + \theta_n^{(2)}$, and we have that under any $\{\psi_n\}$ sequence,

$$\sqrt{n}\theta_n^{(1)} \to (0'_{d_\gamma}, b^{(1)}, 0'_{d_\delta})' \equiv \tau \tag{A.2}$$

and

$$\sqrt{n}(\Theta - \theta_n^{(2)}) \to \tilde{\Lambda} \quad \text{and} \quad \sqrt{n}(\Theta_{\mathsf{H}_0} - \theta_n^{(2)}) \to \tilde{\Lambda}_0 \tag{A.3}$$

where $\Theta_{\mathsf{H}_0} \equiv \{\theta \in \Theta : \gamma = \gamma_0\}$ and $\tilde{\Lambda}$ and $\tilde{\Lambda}_0$ are convex cones (with zero vertex) given respectively by

$$\tilde{\Lambda} = \Lambda_\gamma \times \tilde{\Lambda}_\beta \times \Lambda_\delta \quad \text{and} \quad \tilde{\Lambda}_0 = \{0\}^{d_\gamma} \times \tilde{\Lambda}_\beta \times \Lambda_\delta, \tag{A.4}$$

for $\Lambda_\delta = \mathbb{R}^{d_\delta}$ and $\tilde{\Lambda}_\beta = \tilde{\Lambda}_{\beta,1} \times \cdots \times \tilde{\Lambda}_{\beta,d_\beta}$, with

$$\tilde{\Lambda}_{\beta,i} = \begin{cases} \mathbb{R}_+ & \text{if } b_i < \infty \\ \mathbb{R} & \text{if } b_i = \infty \end{cases}, \quad i = 1, \ldots, d_\beta.$$

As in Andrews (1999), we note that

$$L_n(\theta) = L_n(\theta_n) + \frac{1}{2} Z_n' \Omega_n Z_n - \frac{1}{2} \|\sqrt{n}(\theta - \theta_n) - Z_n\|_{\Omega_n}^2 + R_n(\theta),$$

with

$$\Omega_n \equiv -n^{-1} \frac{\partial^2 L_n(\theta_n)}{\partial\theta \partial\theta'}, \tag{A.5}$$

$$Z_n = \Omega_n^{-1} n^{-1/2} \frac{\partial L_n(\theta_n)}{\partial\theta}, \tag{A.6}$$

$$\|\lambda - Z_n\|_{\Omega_n}^2 = (\lambda - Z_n)' \Omega_n (\lambda - Z_n), \quad \lambda \in \mathbb{R}^{d_\theta}. \tag{A.7}$$

Hence,

$$\begin{aligned} \mathsf{LR}_n &= 2(L_n(\hat{\theta}_n) - L_n(\tilde{\theta}_n)) \\ &= \|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 - \|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 + 2\left(R_n(\hat{\theta}_n) + R_n(\tilde{\theta}_n)\right) \\ &= \|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 - \|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 + o_p(1), \end{aligned}$$

where the last equality follows by Assumptions 1–2. We seek to show that, jointly,

$$\|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 \xrightarrow{d} \inf_{\lambda \in \tilde{\Lambda}_0} \|\lambda - (Z + \tau)\|_{\Omega_0}^2$$

and

$$\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|_{\Omega_n}^2 \xrightarrow{d} \inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z + \tau)\|_{\Omega_0}^2.$$

Given this convergence, the limiting distribution, $\mathcal{L}_\infty(b, b)$, of LR_n is then derived by standard arguments, using the structure of $\tilde{\Lambda}$ and $\tilde{\Lambda}_0$ and that $\Lambda_\delta = \mathbb{R}^{d_\delta}$:

$$\begin{aligned} &\inf_{\lambda \in \tilde{\Lambda}_0} \|\lambda - (Z + \tau)\|_{\Omega_0}^2 - \inf_{\lambda \in \tilde{\Lambda}} \|\lambda - (Z + \tau)\|_{\Omega_0}^2 \\ &= \inf_{\lambda \in \{0\}^{d_\gamma} \times \tilde{\Lambda}_\beta \times \Lambda_\delta} \|\lambda - (Z + \tau)\|_{\Omega_0}^2 - \inf_{\lambda \in \Lambda_\gamma \times \tilde{\Lambda}_\beta \times \Lambda_\delta} \|\lambda - (Z + \tau)\|_{\Omega_0}^2 \\ &= \inf_{\lambda \in \{0\}^{d_\gamma} \times \tilde{\Lambda}_\beta} \|\lambda - H(Z + \tau)\|_{(H\Omega_0^{-1}H')^{-1}}^2 - \inf_{\lambda \in \Lambda_\gamma \times \tilde{\Lambda}_\beta} \|\lambda - H(Z + \tau)\|_{(H\Omega_0^{-1}H')^{-1}}^2 \\ &= \inf_{\lambda \in \{0\}^{d_\gamma} \times \Lambda_\beta(b)} \|\lambda - HZ\|_{(H\Omega_0^{-1}H')^{-1}}^2 - \inf_{\lambda \in \Lambda_\gamma \times \Lambda_\beta(b)} \|\lambda - HZ\|_{(H\Omega_0^{-1}H')^{-1}}^2 \\ &= \mathcal{L}_\infty(b, b). \end{aligned}$$

Let us focus on the convergence of $\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n}$, noting that the convergence of $\|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n}$ follows by similar arguments, and that the convergence holds jointly, as $\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n}$ and $\|\sqrt{n}(\tilde{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n}$ are functions of the same data $\{W_t\}$. It suffices to show the following three properties:

1. With $\hat{\theta}_{q,n}$ satisfying $\|\sqrt{n}(\hat{\theta}_{q,n} - \theta_n) - Z_n\|^2_{\Omega_n} = \inf_{\theta\in\Theta} \|\sqrt{n}(\theta - \theta_n) - Z_n\|^2_{\Omega_n}$, it holds that

$$\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n} = \|\sqrt{n}(\hat{\theta}_{q,n} - \theta_n) - Z_n\|^2_{\Omega_n} + o_p(1).$$

2. It holds that

$$\|\sqrt{n}(\hat{\theta}_{q,n}-\theta_n)-Z_n\|^2_{\Omega_n} = \inf_{\theta\in\Theta} \|\sqrt{n}(\theta-\theta_n)-Z_n\|^2_{\Omega_n} = \inf_{\lambda\in\tilde{\Lambda}} \|\lambda-\sqrt{n}\theta_n^{(1)}-Z_n\|^2_{\Omega_n}+o_p(1).$$

3. It holds that

$$\inf_{\lambda\in\tilde{\Lambda}} \|\lambda - \sqrt{n}\theta_n^{(1)} - Z_n\|^2_{\Omega_n} \xrightarrow{d} \inf_{\lambda\in\tilde{\Lambda}} \|\lambda - (Z+\tau)\|^2_{\Omega_0}.$$

These properties follow from Lemmas A.1–A.3. □

Proof of Lemma 3.1

With $\underline{b} \le b \le \bar{b}$, it holds that $\Lambda_\beta(\underline{b}) \subset \Lambda_\beta(b) \subset \Lambda_\beta(\bar{b})$. This implies that

$$\inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(\underline{b})} Q(\lambda) \ge \inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(b)} Q(\lambda),$$

and

$$\inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(b)} Q(\lambda) \ge \inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(\bar{b})} Q(\lambda).$$

Hence,

$$\begin{aligned}\mathcal{L}_\infty(b,b) &= \inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(b)} Q(\lambda) - \inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(b)} Q(\lambda) \le \inf_{\lambda\in\{0\}^{d_\gamma}\times\Lambda_\beta(\underline{b})} Q(\lambda) - \inf_{\lambda\in\Lambda_\gamma\times\Lambda_\beta(\bar{b})} Q(\lambda)\\ &= \mathcal{L}_\infty(\underline{b},\bar{b}),\end{aligned}$$

as required. □

Proof of Lemma 3.2

By a Taylor-type expansion at θ_n,

$$\sqrt{n}\left(\check{\theta}_n-\theta_n\right)=\left[I_{d_\theta}-\left(\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'}\right]\sqrt{n}(\hat{\theta}_n-\theta_n)$$
$$-\left(n^{-1}\frac{\partial^2 L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}\right)^{-1}n^{-1/2}\frac{\partial L_n(\theta_n)}{\partial\theta}+o_p(1)\xrightarrow{d}N(0,\Omega_0^{-1}\Sigma_0\Omega_0^{-1})$$

where we have used Assumptions 1–4 and Remark 2.1. This proves (3.2). In addition, (3.3) follows directly from Assumptions 1–3 and 5. □

Proof of Theorem 3.3

For the parameter space Ψ, standard subsequencing arguments (e.g., Andrews and Guggenberger, 2009, and McCloskey, 2017) provide that showing

$$\lim_{n\to\infty}\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n\geq\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\leq\alpha \tag{A.8}$$

under all $\{\psi_n\}$ sequences in Ψ satisfying $\psi_n\to\psi_0$ and (2.8) is sufficient for proving the statement of the theorem. Consider any such sequence; then, we have

$$\begin{aligned}
&\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n\geq\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&=\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n\geq\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\geq\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)\right)\\
&\quad+\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n\geq\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)>\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&\quad+\mathbb{P}_{\psi_n}\left(\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)>\mathsf{LR}_n\geq\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&\leq\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n\geq\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)\right) \qquad\qquad (A.9)\\
&\quad+\mathbb{P}_{\psi_n}\left(\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)>\mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right).
\end{aligned}$$

Note the following:

(a) the distribution function of $\mathcal{L}_{\infty,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)$ converges in probability to the distribution function of $\mathcal{L}_\infty(b,b)$ by Assumptions 1–3 and 5 and the continuous mapping theorem;

(b) $\mathsf{LR}_n\xrightarrow{d}\mathcal{L}_\infty(b,b)$ by Lemma 2.1;

(c) for any $b \in [0,\infty]^{d_\beta}$, $\mathcal{L}_\infty(b,b)$ is an absolutely continuous random variable with support $[0,\infty)$.

Therefore, Lemma 5(ii) of Andrews and Guggenberger (2010) implies

$$\mathbb{P}_{\psi_n}\left(\mathsf{LR}_n \geq \mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n)\right) \to \mathbb{P}\left(\mathcal{L}_\infty(b,b) \geq \mathsf{CV}_{1-\alpha+\eta}(b,b)\right) = \alpha-\eta, \quad \text{(A.10)}$$

where $\mathsf{CV}_{1-\alpha+\eta}(b,b)$ denotes the $1-\alpha+\eta$ quantile of $\mathcal{L}_\infty(b,b)$. For $q_{1-\eta}$ equal to the $(1-\eta)$-quantile of $\max_{i=1,\ldots,d_\beta}|Z_{\beta,i}|$ with $Z_\beta \stackrel{d}{=} N(0,\Omega_\beta)$ and $\Omega_\beta = \operatorname{diag}(\Sigma_\beta)^{-1/2}\Sigma_\beta\operatorname{diag}(\Sigma_\beta)^{-1/2}$,

$$\begin{aligned}
&\mathbb{P}_{\psi_n}\left(\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n) > \mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(\mathsf{CV}_{1-\alpha+\eta,n}(\sqrt{n}\beta_n,\sqrt{n}\beta_n) \leq \mathsf{CV}_{1-\alpha+\eta,n}(b_{L,n},b_{U,n})\right)\\
&\leq 1-\mathbb{P}_{\psi_n}\left(b_{L,n} \leq \sqrt{n}\beta_n \leq b_{U,n}\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(\bar{b}_{L,n} \leq \sqrt{n}\beta_n \leq b_{U,n}\right)\\
&\leq 1-\mathbb{P}_{\psi_n}\left(\bar{b}_{L,n} \leq \sqrt{n}\beta_n \leq \bar{b}_{U,n}\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(-\hat{q}_{1-\eta,n}\operatorname{diagv}(\hat{\Sigma}_{\beta,n})^{1/2} \leq \sqrt{n}(\bar{\beta}_n-\beta_n) \leq \hat{q}_{1-\eta,n}\operatorname{diagv}(\hat{\Sigma}_{\beta,n})^{1/2})\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(-\hat{q}_{1-\eta,n} \leq \frac{\sqrt{n}(\bar{\beta}_{n,i}-\beta_{n,i})}{\sqrt{\hat{\Sigma}_{\beta,n,ii}}} \leq \hat{q}_{1-\eta,n} \text{ for all } i=1,\ldots,d_\beta\right)\\
&= 1-\mathbb{P}_{\psi_n}\left(\max_{i=1,\ldots,d_\beta}\left|\frac{\sqrt{n}(\bar{\beta}_{n,i}-\beta_{n,i})}{\sqrt{\hat{\Sigma}_{\beta,n,ii}}}\right| \leq \hat{q}_{1-\eta,n}\right)\\
&\to 1-\mathbb{P}\left(\max_{i=1,\ldots,d_\beta}|Z_{\beta,i}| \leq q_{1-\eta}\right) = \eta, \qquad \text{(A.11)}
\end{aligned}$$

where the inequalities inside of probabilities are evaluated element-wise across vectors, the first inequality follows from Lemma 3.1, the second equality follows from $\beta_n \geq 0$ and the convergence follows from Assumption 6 and Lemma 5(ii) of Andrews and Guggenberger (2010).

Together, (A.9)–(A.11) imply (A.8), and therefore the statement of the theorem. □

A.2 Technical lemmas for proving main results

Lemma A.1 *Let $\hat{\theta}_{q,n}$ satisfy $\|\sqrt{n}(\hat{\theta}_{q,n}-\theta_n)-Z_n\|^2_{\Omega_n} = \inf_{\theta\in\Theta}\|\sqrt{n}(\theta-\theta_n)-Z_n\|^2_{\Omega_n}$, with Ω_n given by (A.5) and Z_n given by (A.6). Under Assumptions 1-4 and any sequence $\{\psi_n\}$*

satisfying $\psi_n \to \psi_0$ *and (2.8),*

$$\|\sqrt{n}(\hat{\theta}_n - \theta_n) - Z_n\|^2_{\Omega_n} = \|\sqrt{n}(\hat{\theta}_{q,n} - \theta_n) - Z_n\|^2_{\Omega_n} + o_p(1).$$

PROOF. The result follows directly from arguments given in Andrews (1999, proof of Theorem 2) with θ_0 replaced by θ_n. □

LEMMA A.2 *Under Assumptions 3-4 and any sequence* $\{\psi_n\}$ *satisfying* $\psi_n \to \psi_0$ *and (2.8), it holds that*

$$\inf_{\theta\in\Theta} \|\sqrt{n}(\theta - \theta_n) - Z_n\|^2_{\Omega_n} = \inf_{\lambda\in\tilde{\Lambda}} \|\lambda - \sqrt{n}\theta_n^{(1)} - Z_n\|^2_{\Omega_n} + o_p(1),$$

where Ω_n *given by (A.5),* Z_n *given by (A.6),* $\tilde{\Lambda}$ *is given by (A.4) and* $\theta_n^{(1)}$ *is given by (A.1).*

PROOF. By definition, using (A.1),

$$\|\sqrt{n}(\theta - \theta_n) - Z_n\|^2_{\Omega_n} = \|\sqrt{n}(\theta - \theta_n^{(2)}) - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n}.$$

Hence,

$$\begin{aligned}\inf_{\theta\in\Theta} \|\sqrt{n}(\theta - \theta_n) - Z_n\|^2_{\Omega_n} &= \inf_{\theta\in\Theta} \|\sqrt{n}(\theta - \theta_n^{(2)}) - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n} \\ &= \inf_{\lambda\in\sqrt{n}(\Theta-\theta_n^{(2)})} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n}.\end{aligned}$$

By Assumptions 3-4 and (A.2), we have that $Z_n + \sqrt{n}\theta_n^{(1)} = O_p(1)$. Hence, the result now follows using the fact that $\sqrt{n}(\Theta - \theta_n^{(2)})$ contains zero for all $n \geq 1$ and Silvapulle and Sen (2005, Corollary 4.7.5.1 and the comments on p. 194). □

LEMMA A.3 *Under Assumptions 3-4 and any sequence* $\{\psi_n\}$ *satisfying* $\psi_n \to \psi_0$ *and (2.8), it holds that*

$$\inf_{\lambda\in\tilde{\Lambda}} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n} \xrightarrow{d} \inf_{\lambda\in\tilde{\Lambda}} \|\lambda - (Z + \tau)\|^2_{\Omega_0},$$

where Ω_n *given by (A.5),* Z_n *given by (A.6),* $\tilde{\Lambda}$ *is given by (A.4),* $\theta_n^{(1)}$ *is given by (A.1), and* τ *is defined in (A.2).*

PROOF. Since $(Z_n + \sqrt{n}\theta_n^{(1)}) = O_p(1)$ and $\Omega_n = \Omega_0 + o_p(1)$ by Assumptions 3-4, we have by Silvapulle and Sen (2005, Lemma 4.10.2.2) that

$$\inf_{\lambda\in\tilde{\Lambda}} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_n} = \inf_{\lambda\in\tilde{\Lambda}} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_0} + o_p(1).$$

By Silvapulle and Sen (2005, Corollary 4.7.5.2) and the fact that $Z_n + \sqrt{n}\theta_n^{(1)} \xrightarrow{d} Z + \tau$, we have that $\inf_{\lambda\in\tilde{\Lambda}} \|\lambda - (Z_n + \sqrt{n}\theta_n^{(1)})\|^2_{\Omega_0} \xrightarrow{d} \inf_{\lambda\in\tilde{\Lambda}} \|\lambda - (Z + \tau)\|^2_{\Omega_0}$. □

A.3 Proofs and lemmas related to linear regression example

Proof of Proposition 2.1

Starting with Assumption 1, note that Assumptions LinIID 1.1–3 and a weak LLN for row-wise i.i.d. random variables imply that $\|S_{xx} - \mathbb{E}_{\psi_n}[x_t x_t']\| \xrightarrow{p} 0$ and $\|S_{x\varepsilon} - \mathbb{E}_{\psi_n}[x_t\varepsilon_t]\| \xrightarrow{p} 0$ (all convergence statements that follow are thus understood to be under any sequence $\{\psi_n\}$ satisfying $\psi_n \to \psi_0$ and (2.8)). Since x_t under ψ_n converges in distribution to x_t under ψ_0 and $\max_j E_\psi[|x_{t,j}|^{2+\nu}] \leq c$ for all $\psi \in \Psi$, we have that $\mathbb{E}_{\psi_n}[x_t x_t'] \to \mathbb{E}_{\psi_0}[x_t x_t'] = \Omega_0$, such that $S_{xx} \xrightarrow{p} \Omega_0$. Likewise, $\mathbb{E}_\psi[x_t\varepsilon_t] = 0$ for all $\psi \in \Psi$, such that $S_{x\varepsilon} \xrightarrow{p} 0$. Clearly, these convergence properties imply that S_{xx} is invertible with probability approaching one, such that

$$\hat{\theta}_{LS} - \theta_n = S_{xx}^{-1} S_{x\varepsilon} = o_p(1), \tag{A.12}$$

and

$$\|\hat{\theta}_{LS} - \theta_n\|^2_{S_{xx}} = (\hat{\theta}_{LS} - \theta_n)' S_{xx} (\hat{\theta}_{LS} - \theta_n) = o_p(1).$$

We then have by the triangle inequality

$$\|\hat{\theta}_n - \theta_n\|_{S_{xx}} \leq \|\hat{\theta}_n - \hat{\theta}_{LS}\|_{S_{xx}} + \|\hat{\theta}_{LS} - \theta_n\|_{S_{xx}} \leq 2\|\hat{\theta}_{LS} - \theta_n\|_{S_{xx}},$$

where the second equality follows by noting that $\theta_n \in \Theta$ and $\|\hat{\theta}_n - \hat{\theta}_{LS}\|_{S_{xx}} = \min_{\theta\in\Theta} \|\theta - \hat{\theta}_{LS}\|_{S_{xx}}$. We conclude that $\|\hat{\theta}_n - \theta_n\|_{S_{xx}} = o_p(1)$, and hence that $\hat{\theta}_n - \theta_n = o_p(1)$. By similar arguments we have that $\tilde{\theta}_n - \theta_n = o_p(1)$.

Moving now to Assumption 2, 1. clearly holds by the definition of $L_n(\cdot)$ and 2. holds trivially since $-n^{-1}\partial^2 L_n(\theta)/\partial\theta\partial\theta' = S_{xx}$ in this example. Assumption 3 also holds by the fact that $-n^{-1}\partial^2 L_n(\theta)/\partial\theta\partial\theta' = S_{xx}$ and that $S_{xx} \xrightarrow{p} \Omega_0$. Finally, Assumption 4 holds by Assumptions LinIID 1.1–3, the Cramér Wold device and a Liapounov central limit theorem (CLT) for row-wise i.i.d. random variables since $n^{-1/2}\partial L_n(\theta_n)/\partial\theta = n^{1/2}S_{x\varepsilon}$ in this example. $\square$

Proof of Proposition 3.1

The proof that Assumption 5 holds is nearly identical to those of White (1980, proof of Theorem 1). For Assumption 6, note that under any sequence $\{\psi_n\}$ satisfying $\psi_n \to \psi_0$ and (2.8),

$$\sqrt{n}(\hat{\theta}_{LS} - \theta_n) = S_{xx}^{-1}\sqrt{n}S_{x\varepsilon} \xrightarrow{d} N(0, \Omega_0^{-1}\Sigma_0\Omega_0^{-1})$$

by the continuous mapping theorem and the fact that Assumptions 3–4 hold. Furthermore, given that Assumptions 3 and 5 hold, the continuous mapping theorem implies $\hat{\Sigma}_{\beta,n} \xrightarrow{p} \Sigma_\beta$. $\square$

Lemma A.4 *Suppose that S_{xx} is invertible. For any set $\Lambda \subset \mathbb{R}^{1+d_\beta}$,*

$$\arg\inf_{\theta\in\Lambda}\sum_{t=1}^{n}(y_t - x_t'\theta)^2 = \arg\inf_{\theta\in\Lambda}(\theta - \hat{\theta}_{LS})'S_{xx}(\theta - \hat{\theta}_{LS}).$$

Proof. After noting that for any θ,

$$\begin{aligned}\sum_{t=1}^{n}(y_t - x_t'\theta)^2 &= \sum_{t=1}^{n}(y_t - x_t'\hat{\theta}_{LS} - x_t'(\theta - \hat{\theta}_{LS}))^2 \\ &= \sum_{t=1}^{n}(y_t - x_t'\hat{\theta}_{LS})^2 + \sum_{t=1}^{n}(x_t'(\theta - \hat{\theta}_{LS}))^2 \\ &\quad - 2(\theta - \hat{\theta}_{LS})'\underbrace{\sum_{t=1}^{n}x_t(y_t - x_t'\hat{\theta}_{LS})}_{=0} \\ &= \sum_{t=1}^{n}(y_t - x_t'\hat{\theta}_{LS})^2 + n(\theta - \hat{\theta}_{LS})'S_{xx}(\theta - \hat{\theta}_{LS}),\end{aligned}$$

the result follows immediately. $\square$

A.4 Proofs related to the ARCH example

Below we prove that Assumptions 1-6 hold under Assumptions ARCH 1-4. Throughout, we make use of

$$W_t = (y_t^2, F_{t-1}')' \in \mathcal{W} = \mathbb{R}_+ \times \{1\} \times \mathbb{R}_+^{d_\beta + d_\gamma}.$$

Proof that Assumption 1 holds

Under under any sequence $\{\psi_n\}$ satisfying $\psi_n \to \psi_0$ and (2.8), and using the compactness of Θ, to prove the convergence of $\hat{\theta}_n$, it suffices to show that

$$\sup_{\theta \in \Theta} |n^{-1} L_n(\theta) - \mathcal{L}(\theta)| \xrightarrow{p} 0, \tag{A.13}$$

with

$$\mathcal{L}(\theta) = -\frac{1}{2}\mathbb{E}_{\psi_0}\left[\log \sigma_t^2(\theta) + \frac{y_t^2}{\sigma_t^2(\theta)}\right] \tag{A.14}$$

and

$$\mathcal{L}(\theta) \leq \mathcal{L}(\theta_0) \text{ for any } \theta \in \Theta \text{ with equality if and only if } \theta = \theta_0. \tag{A.15}$$

We start out by showing that (A.13) by applying Lemma 11.3 of Andrews and Cheng (2013b)[2]. Recall that

$$l_t(\theta) = -\frac{1}{2}\left(\log \sigma_t^2(\theta) + \frac{y_t^2}{\sigma_t^2(\theta)}\right) = -\frac{1}{2}\left(\log g(F_{t-1}, \theta) + \frac{W_{t,1}}{g(F_{t-1}, \theta)}\right),$$

with $g(F_{t-1}, \theta) := F_{t-1}'\theta$ and $W_{t,1} = y_t^2$, the first entry of W_t. For any $w \in \mathcal{W}$, let w_1 denote the first entry of w and w_2 the column vector of the remaining entries, that is, $w = (w_1, w_2')'$. Let

$$s(w, \theta) = \log g(w, \theta) + \frac{w_1}{g(w_2, \theta)} = \log (w_2'\theta) + \frac{w_1}{w_2'\theta}.$$

For any $\theta_1, \theta_2 \in \Theta$, a mean value expansion gives

$$\log (w_2'\theta_1) = \log (w_2'\theta_2) + \frac{1}{w_2'\theta^*} w_2'(\theta_1 - \theta_2),$$

with $\theta^* \in \Theta$ between θ_1 and θ_2. Likewise,

$$\frac{w_1}{w_2'\theta_1} = \frac{w_1}{w_2'\theta_2} - \frac{w_1}{(w_2'\theta^{**})^2} w_2'(\theta_1 - \theta_2),$$

[2] A careful inspection of the proof of that lemma shows that only strong mixing conditions as the ones stated in Assumption ARCH 1 are needed. In particular, the proof makes use of a weak LLN for triangular arrays of strongly mixing processes, which does not impose any rate of decay on the the mixing coefficients.

with $\theta^{**} \in \Theta$ between θ_1 and θ_2. It holds that $w_2'\theta \geq \delta_L$ uniformly on $\mathcal{W} \times \Theta$. Consequently, for all $\theta_1, \theta_2 \in \Theta$

$$\begin{aligned}|s(w,\theta_1) - s(w,\theta_2)| &= \left| \frac{1}{w_2'\theta^*} w_2'(\theta_1 - \theta_2) - \frac{w_1}{(w_2'\theta^{**})^2} w_2'(\theta_1 - \theta_2) \right| \\ &\leq \left(\delta_L^{-1} + \delta_L^{-2} w_1\right) \|w_2\| \, \|\theta_1 - \theta_2\| \,.\end{aligned}$$

With $M_1(w) := \left(\delta_L^{-1} + \delta_L^{-2} w_1\right) \|w_2\|$, we conclude that for any $\eta > 0$

$$|s(w,\theta_1) - s(w,\theta_2)| \leq M_1(w)\eta \tag{A.16}$$

for all $\theta_1, \theta_2 \in \Theta$ and $w \in \mathcal{W}$ with $\|\theta_1 - \theta_2\| < \eta$. By Assumption ARCH 4, it holds that

$$\begin{aligned}\mathbb{E}_\psi[M_1(W_t)] &= \mathbb{E}_\psi\left[\left(\delta_L^{-1} + \delta_L^{-2} y_t^2\right) \|F_{t-1}\|\right] \\ &\leq \delta_L^{-1}\mathbb{E}_\psi\left[\|F_{t-1}\|\right] + \delta_L^{-2}\mathbb{E}_\psi\left[y_t^2 \|F_{t-1}\|\right] \leq \tilde{c}\end{aligned}$$

for some constant $\tilde{c} \in (0,\infty)$ for all $\psi \in \Psi$. Moreover, we have that for $\theta \in \Theta$

$$|s(W_t,\theta)| \leq |\log(\delta_L)| + \|F_{t-1}\| d_\theta(\delta_U + \beta_U + \gamma_U) + \frac{y_t^2}{\delta_L},$$

so using Assumption ARCH 4 again, we have that

$$\mathbb{E}_\psi[\sup_{\theta\in\Theta} |s(W_t,\theta)|^{1+\nu}] \leq \tilde{c}$$

for some constants $\nu, \tilde{c} \in (0,\infty)$ for all $\psi \in \Psi$. We conclude that

$$\mathbb{E}_\psi[\sup_{\theta\in\Theta} |s(W_t,\theta)|^{1+\nu}] + \mathbb{E}_\psi[M_1(W_t)] \leq \bar{C}, \tag{A.17}$$

for some constants $\nu, \tilde{c} \in (0,\infty)$ for all $\psi \in \Psi$. Using the fact that $l_t(\theta) = -s(W_t,\theta)/2$ together with (A.16) and (A.17), we have that (A.13) holds by Lemma 11.3 of Andrews and Cheng (2013b). Condition (A.15) holds by standard arguments and Assumption 2. The properties (A.13)-(A.15) imply the convergence of $\hat{\theta}_n$. By identical arguments (under H_0) we can prove that $\tilde{\theta}_n$ converges, replacing Θ by Θ_{H_0}. □

Proof that Assumption 2 holds

Note that

$$\sup_{\theta\in\Theta:\|\theta-\theta_n\|\leq\epsilon_n} \left\| n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'} - n^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'} \right\|$$

$$\leq 2 \sup_{\theta \in \Theta} \left\| n^{-1} \frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'} - \mathbb{E}_{\psi_0} \left[\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right] \right\|$$
$$+ \sup_{\theta \in \Theta : \|\theta - \theta_n\| \leq \epsilon_n} \left\| \mathbb{E}_{\psi_0} \left[\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right] - \mathbb{E}_{\psi_0} \left[\frac{\partial^2 l_t(\theta_n)}{\partial \theta \partial \theta'} \right] \right\|.$$

Hence Assumption 2 holds provided

$$\sup_{\theta \in \Theta} \left\| n^{-1} \frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'} - \mathbb{E}_{\psi_0} \left[\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} \right] \right\| = o_p(1), \tag{A.18}$$

and that $\mathbb{E}_{\psi_0}[\partial^2 l_t(\theta) / \partial \theta \partial \theta']$ is continuous. Both conditions are shown by an application of Lemma 11.3 of Andrews and Cheng (2013b), and the proof follows closely the arguments given in the previous proof. Note initially, that (A.18) holds provided that for any $i, j = 1, \ldots, d_\theta$

$$\sup_{\theta \in \Theta} \left| n^{-1} \frac{\partial^2 L_n(\theta)}{\partial \theta_i \partial \theta_j} - \mathbb{E}_{\psi_0} \left[\frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} \right] \right| = o_p(1). \tag{A.19}$$

It holds that for any $i, j = 1, \ldots, d_\theta$,

$$\frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} = -\frac{1}{2} \left[2 \frac{y_t^2}{\sigma_t^6(\theta)} - \frac{1}{\sigma_t^4(\theta)} \right] \left(\frac{\partial \sigma_t^2(\theta)}{\partial \theta_i} \right) \left(\frac{\partial \sigma_t^2(\theta)}{\partial \theta_j} \right)$$
$$= -\frac{1}{2} \left[2 \frac{y_t^2}{(F_{t-1}' \theta)^3} - \frac{1}{(F_{t-1}' \theta)^2} \right] F_{t-1,i} F_{t-1,j} = s_{ij}(W_t, \theta),$$

with

$$s_{ij}(w, \theta) = -\frac{1}{2} \left[2 \frac{w_1}{(w_2' \theta)^3} - \frac{1}{(w_2' \theta)^2} \right] w_{2,i} w_{2,j}, \quad w = (w_1, w_2')' \in \mathcal{W}.$$

For any $\theta_1, \theta_2 \in \Theta$, a mean value expansion gives that

$$\frac{w_1}{(w_2' \theta_1)^3} = \frac{w_1}{(w_2' \theta_2)^3} - 3 \frac{w_1}{(w_2' \theta^*)^4} w_2' (\theta_1 - \theta_2),$$

and

$$\frac{1}{(w_2' \theta_1)^2} = \frac{1}{(w_2' \theta_2)^2} - 2 \frac{1}{(w_2' \theta^{**})^3} w_2' (\theta_1 - \theta_2)$$

with $\theta^*, \theta^{**} \in \Theta$ between θ_1 and θ_2. Consequently, for any $\theta_1, \theta_2 \in \Theta$,

$$s_{ij}(w, \theta_1) - s_{ij}(w, \theta_2) = -\frac{1}{2} \left[2 \frac{w_1}{(w_2' \theta_1)^3} - \frac{1}{(w_2' \theta_1)^2} - \left(2 \frac{w_1}{(w_2' \theta_2)^3} - \frac{1}{(w_2' \theta_2)^2} \right) \right] w_{2,i} w_{2,j}$$
$$= \left[\frac{3 w_1}{(w_2' \theta^*)^4} - \frac{1}{(w_2' \theta^{**})^3} \right] w_{2,i} w_{2,j} w_2' (\theta_1 - \theta_2),$$

such that

$$|s_{ij}(w,\theta_1) - s_{ij}(w,\theta_2)| \leq \underbrace{\left(\frac{3w_1}{\delta_L^4} + \frac{1}{\delta_L^3}\right) w_{2,i}w_{2,j}\|w_2\|}_{:=M_{ij}(w)}\|\theta_1 - \theta_2\|.$$

We conclude that for any $\eta > 0$

$$|s_{ij}(w,\theta_1) - s_{ij}(w,\theta_2)| \leq M_{ij}(w)\eta \tag{A.20}$$

for all $\theta_1, \theta_2 \in \Theta$ and $w \in \mathcal{W}$ with $\|\theta_1 - \theta_2\| < \eta$. By Assumption ARCH 4, it holds that

$$\begin{aligned}\mathbb{E}_\psi[M_{ij}(W_t)] &= \mathbb{E}_\psi\left[\left(\frac{3y_t^2}{\delta_L^4} + \frac{1}{\delta_L^3}\right) F_{t-1,i}F_{t-1,j}\|F_{t-1}\|\right] \\ &\leq \frac{3}{\delta_L^4}\mathbb{E}_\psi\left[y_t^2\|F_{t-1}\|^3\right] + \delta_L^{-3}\mathbb{E}_\psi\left[\|F_{t-1}\|^3\right] \leq \tilde{c}\end{aligned}$$

for some constant $\tilde{c} \in (0,\infty)$ for all $\psi \in \Psi$. Moreover, we have that for $\theta \in \Theta$

$$|s_{ij}(W_t,\theta)| \leq \frac{1}{2}\left[2\frac{y_t^2}{\delta_L^3} + \frac{1}{\delta_L^2}\right] F_{t-1,i}F_{t-1,j},$$

so applying Assumption ARCH 4 again, we have that

$$\mathbb{E}_\psi \sup_{\theta\in\Theta}[|s_{ij}(W_t,\theta)|^{1+\nu}] \leq c^*$$

for some constants $\nu, c^* \in (0,\infty)$ for all $\psi \in \Psi$. We conclude that for any $i,j = 1,\ldots,d_\theta$,

$$\mathbb{E}_\psi \sup_{\theta\in\Theta}[|s_{ij}(W_t,\theta)|^{1+\nu}] + \mathbb{E}_\psi[M_{ij}(W_t)] \leq c, \tag{A.21}$$

for some constant $c \in (0,\infty)$ for all $\psi \in \Psi$. Using (A.20) and (A.21) together with Lemma 11.3 of Andrews and Cheng (2013b), we have that (A.19) holds and, hence, that (A.18) holds. Moreover, this lemma ensures that $\mathbb{E}_{\psi_0}[\partial^2 l_t(\theta)/\partial\theta_i\partial\theta_j]$ is uniformly continuous on Θ for all $\psi_0 \in \Psi$. □

Proof that Assumption 3 holds

Recall that $\Omega_0 = -\mathbb{E}_{\psi_0}[\partial^2 l_t(\theta_0)/\partial\theta\partial\theta']$ and note that the matrix is positive definite for all $\psi_0 \in \Psi$ under Assumptions ARCH 1-4, by standard arguments. It holds that

$$\begin{aligned}\left\|-n^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'} - \Omega_0\right\| &= \left\|n^{-1}\frac{\partial^2 L_n(\theta_n)}{\partial\theta\partial\theta'} - \mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right]\right\| \\ &\leq \sup_{\theta\in\Theta}\left\|n^{-1}\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'} - \mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\right]\right\|\end{aligned}$$

$$+\left\|\mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right]-\mathbb{E}_{\psi_0}\left[\frac{\partial^2 l_t(\theta_n)}{\partial\theta\partial\theta'}\right]\right\|,$$

where the first term is $o_p(1)$ for all $\psi_0 \in \Psi$, by the arguments given in the proof of Assumption *2*. Moreover, from that proof, it holds that the second term is $o(1)$ for all $\psi_0 \in \Psi$. □

Proof that Assumption 4 holds

For a given ψ_n, the (scaled) score is given by

$$S_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial l_t(\theta_n)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} -\frac{1}{2}(\varepsilon_t^2-1)\frac{1}{\sigma_t^2(\theta_n)}F_{t-1},$$

with F_t being $\mathcal{F}_{t,n}$-measurable. For any non-zero constant vector $k \in \mathbb{R}^{d_\theta}$, let $s_t = -k'F_{t-1}\sigma_t^{-2}(\theta_n)(\varepsilon_t^2-1)/2$, such that $k'S_n = n^{-1/2}\sum_{t=1}^{n} s_t$, and note that by Assumption ARCH 3, $\mathbb{E}_{\psi_n}[s_t|\mathcal{F}_{t-1,n}] = 0$ almost surely. The result follows by an application of the Lindeberg CLT for martingale difference arrays combined with an application of the Cramér-Wold Theorem. Note that by Assumptions ARCH 3 and ARCH 4,

$$\begin{aligned}\mathbb{E}_{\psi_n}[s_t^2] &= \mathbb{E}_{\psi_n}[(\varepsilon_t^2-1)^2/4]k'\mathbb{E}_{\psi_n}[\sigma_t^{-4}(\theta_n)F_{t-1}F_{t-1}']k \\ &= (\kappa/4)k'\mathbb{E}_{\psi_n}[\sigma_t^{-4}(\theta_n)F_{t-1}F_{t-1}']k.\end{aligned}$$

The convergence of parameters under the drifting sequence induces convergence in distribution of (F_{t-1},θ_n) under ψ_n to (F_{t-1},θ_0) under ψ_0. Consequently, by the continuous mapping theorem $\sigma_t^{-4}(\theta_n)F_{t-1}F_{t-1}' = (\theta_n'F_{t-1})^{-2}F_{t-1}F_{t-1}'$ converges in distribution to $(\theta_0'F_{t-1})^{-2}F_{t-1}F_{t-1}'$ under ψ_0. Assumption ARCH 4 implies that $\|F_{t-1}\sigma_t^{-2}(\theta_n)\|^2 \le \delta_L^{-2}\|F_{t-1}\|^2$ is uniformly integrable, and consequently, we have that $\mathbb{E}_{\psi_n}[\sigma_t^{-4}(\theta_n)F_{t-1}F_{t-1}'] \to \mathbb{E}_{\psi_0}[\sigma_t^{-4}(\theta_0)F_{t-1}F_{t-1}']$. Hence,

$$\mathbb{E}_{\psi_n}[s_t^2] \to k'\Sigma_0 k,$$

with

$$\Sigma_0 := (\kappa/4)\,\mathbb{E}_{\psi_0}[\sigma_t^{-4}(\theta_0)F_{t-1}F_{t-1}'].$$

The matrix Σ_0 is positive definite by standard arguments and Assumption ARCH 2. It remains to show that for any constant $\epsilon > 0$,

$$\frac{1}{n}\sum_{t=1}^{n} s_t^2 \mathbb{I}(s_t^2 > \sqrt{n}\epsilon) \xrightarrow{p} 0, \tag{A.22}$$

and

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[s_t^2|\mathcal{F}_{t-1,n}\right] - \mathbb{E}_{\psi_n}[s_t^2] \xrightarrow{p} 0. \tag{A.23}$$

To show (A.22), note that by Assumptions ARCH 3 and ARCH 4 and Minkowski's inequality there exist constants a $v, c \in (0, \infty)$ (with c depending on k) such that for any $\eta > 0$,

$$\begin{aligned}
&\mathbb{P}_{\psi_n}\left(\frac{1}{n}\sum_{t=1}^{n} s_t^2 \mathbb{I}(s_t^2 > \sqrt{n}\epsilon) > \eta\right) \leq \eta^{-1}\mathbb{E}_{\psi_n}\left[s_t^2 \mathbb{I}(s_t^2 > \sqrt{n}\epsilon)\right] \\
&\leq \frac{1}{\eta\left(\sqrt{n}\epsilon\right)^{v}}\mathbb{E}_{\psi_n}\left[|s_t|^{2+v}\right] = \frac{1}{\eta\left(\sqrt{n}\epsilon\right)^{v}}\mathbb{E}_{\psi_n}\left[|(\varepsilon_t^2 - 1)\sigma_t^{-2}(\theta_n)k'F_{t-1}/2|^{2+v}\right] \\
&= \frac{1}{2^{2+v}\eta\left(\sqrt{n}\epsilon\right)^{v}}\mathbb{E}_{\psi_n}\left[|\varepsilon_t^2 - 1|^{2+v}\right]\mathbb{E}_{\psi_n}\left[|\sigma_t^{-2}(\theta_n)k'F_{t-1}|^{2+v}\right] \\
&\leq \frac{1}{2^{2+v}\eta\left(\sqrt{n}\epsilon\right)^{v}}\mathbb{E}_{\psi_n}\left[|\varepsilon_t^2 - 1|^{2+v}\right]\delta_L^{-(2+v)}\left(\sum_{i=1}^{d_\theta}|k_i|(\mathbb{E}_{\psi_n}\left[\|F_{t-1}\|^{2+v}\right])^{1/(2+v)}\right)^{2+v} \\
&\leq \frac{1}{2^{2+v}\eta\left(\sqrt{n}\epsilon\right)^{v}}c \to 0,
\end{aligned}$$

and we conclude that (A.22) holds. The convergence in (A.23) follows by an application of the (weak) LLN for row-wise stationary and strongly mixing triangular arrays (Andrews, 1988, p. 462), using that $\mathbb{E}_{\psi_n}[s_t^2|\mathcal{F}_{t-1,n}] = (\kappa/4)(\sigma_t^{-2}(\theta_n)k'F_{t-1})^2$ is uniformly integrable under Assumption ARCH 4. □

Proof that Assumption 5 holds

First, note that by Assumptions 1–3, $\hat{\Omega}_n \xrightarrow{p} \Omega_0$. Consequently, it remains to show that

$$\hat{\kappa}_n \xrightarrow{p} \kappa. \tag{A.24}$$

We have that

$$\hat{\kappa}_n = n^{-1}\sum_{t=1}^{n}\left(\frac{y_t^4}{(\hat{\theta}_n' F_{t-1})^2} - 1\right).$$

Using the same notation as before, let $f : \mathcal{W} \times \times \to\mathbb{R}$ be given by

$$f(w, \theta) = \frac{w_1^2}{(\theta' w_2)^2},$$

such that $\hat{\kappa}_n = n^{-1} \sum_{t=1}^{n} f(W_t, \hat{\theta}_n)$. With $\theta_1, \theta_2 \in \Theta$, a mean-value expansion gives that

$$f(w, \theta_1) - f(w, \theta_2) = -2 \frac{w_1^2}{(\theta^{*\prime} w_2)^3} w_2'(\theta_1 - \theta_2).$$

It holds that

$$| - 2 \frac{w_1^2}{(\theta^{*\prime} w_2)^3} w_2'(\theta_1 - \theta_2)| \leq \frac{2w_1^2}{\delta_L^3} \|w_2\| \|\theta_1 - \theta_2\|$$

such that with $M : \mathcal{W} \to\mathbb{R}$ given by

$$M(w) = \frac{2w_1^2}{\delta_L^3} \|w_2\|,$$

$$|f(w, \theta_1) - f(w, \theta_2)| \leq M(w) \|\theta_1 - \theta_2\|.$$

Consequently, for any $\eta > 0$

$$|f(w, \theta_1) - f(w, \theta_2)| \leq M(w)\eta \tag{A.25}$$

for all $\theta_1, \theta_2 \in \Theta$ and $w \in \mathcal{W}$ with $\|\theta_1 - \theta_2\| < \eta$. By Assumption ARCH 4,

$$\mathbb{E}_\psi[M(W_t)] = \frac{2}{\delta_L^3} \mathbb{E}_\psi \left[y_t^2 \|F_{t-1}\| \right] \leq \tilde{c},$$

for some constant $\tilde{c} \in (0, \infty)$ for all $\psi \in \Psi$. Likewise, noting that for any $\theta \in \Theta$, $|f(W_t, \theta)| \leq y_t^4 / \delta_L^2$, we have by Assumption ARCH 4,

$$\mathbb{E}_\psi \left[\sup_{\theta \in \Theta} |f(W_t, \theta)|^{1+\varepsilon} \right] \leq c^*$$

for some constants $\varepsilon, c^* \in (0, \infty)$ for all $\psi \in \Psi$. Consequently,

$$\mathbb{E}_\psi \left[\sup_{\theta \in \Theta} |f(W_t, \theta)|^{1+\nu} \right] + \mathbb{E}_\psi[M(W_t)] \leq \bar{c}, \tag{A.26}$$

for some constant $\bar{c} \in (0, \infty)$ for all $\psi \in \Psi$. Using (A.25) and (A.26) together with Lemma 11.3 of Andrews and Cheng (2013b), we have that

$$\sup_{\theta \in \Theta} \left| n^{-1} \sum_{t=1}^{n} f(W_t, \theta) - \mathbb{E}_{\psi_0}[f(W_t, \theta)] \right| = o_p(1)$$

and that $\mathbb{E}_{\psi_0}[f(W_t,\theta)]$ is uniformly continuous on Θ for all $\psi_0 \in \Psi$. Using that $\hat{\theta}_n - \theta_0 = o_p(1)$, we then have that

$$n^{-1}\sum_{t=1}^{n} f(W_t,\hat{\theta}_n) - \mathbb{E}_{\psi_0}[f(W_t,\theta_0)] = o_p(1),$$

or, equivalently, (A.24) holds. □

PROOF THAT ASSUMPTION 6 HOLDS

As $\check{\beta}_n$ is given in terms of the Newton-Raphson estimator, Assumption 6 holds by Lemma 3.2. □

B ADDITIONAL NUMERICAL RESULTS FOR ARCH

In this section we provide additional simulation results, complementing the findings in Section 4.2. The data generating process for the simulations is given by

$$y_t = \sigma_t z_t, \quad t = 1,\dots,n,$$
$$\sigma_t^2 = \delta_1 + \delta_2 y_{t-1}^2 + \gamma x_{t-1,1} + \beta_1 x_{t-1,2} + \beta_2 x_{t-1,3} + \beta_3 x_{t-1,4},$$

where $\delta_1 > 0$, $\delta_2, \gamma, \beta_1, \beta_2, \beta_3 \geq 0$, and $\{z_t\}_{t=1,\dots,n}$ is an i.i.d. process with $z_t \sim N(0,1)$. We seek to test the hypothesis

$$\mathsf{H}_0 : \gamma = 0, \tag{A.1}$$

against $\gamma > 0$.

In terms of the covariates, we let

$$x_{i,t} = \frac{\tilde{X}_{i,t}}{\mathbb{E}[\tilde{X}_{i,t}]}, \quad i = 1,\dots,4, \quad t = 1,\dots,T,$$

where

$$\tilde{X}_{i,t} = F_i^{-1}(U_{i,t}), \quad i = 1,\dots,4,$$

with $F_i^{-1}(\cdot)$ the inverse distribution function of $\Gamma(a_i, b)$ for $i = 1,2,3$, $b = 10$ and $a_1 = 3, a_2 = 5, a_3 = 10$, and $F_4^{-1}(\cdot)$ is the inverse distribution function of χ_5^2. The correlated uniform variables $U_{i,t} = \Phi(Z_{i,t})$ for $i = 1,\dots,4$, where $(Z_t)_{t=0}^{T}$ is an i.i.d. process with $Z_t = (Z_{1,t},\dots,Z_{4,t})' \sim N_4(0,\Sigma)$ and Σ a positive definite correlation matrix.

For the experiment we assume that it is known to the researcher that the true value of the ARCH coefficient δ_2 is not near its boundary of zero, so that the only parameters that potentially cause a discontinuity in the null distribution are $\beta_1, \beta_2, \beta_3$. We report the rejection frequencies for $n = 5000$ observations, parameter values $\beta_1 = \beta_2 = 0$, $\gamma, \beta_3 \in \{0, 0.01, 0.05, 0.1, 0.25\}$ and

$$\Sigma = \begin{bmatrix} 1 & & & \\ -0.75 & 1 & & \\ -2/3 & 0.4 & 1 & \\ -0.1 & 0.15 & 0.35 & 1 \end{bmatrix},$$

and compare with the standard LR as well as the CLR test.

Table 8 contains rejection frequencies for different values of β_3 and γ.

[Table 8 around here]

References

Andrews, D.W.K. (1988): "Laws of large numbers for dependent non-identically distributed random variables", *Econometric Theory*, vol. 4, 458–467.

Andrews, D.W.K. (1999): "Estimation when a parameter is on a boundary", *Econometrica*, vol. 67, 1341–1383.

Andrews, D.W.K. and Cheng, X. (2013),"Maximum likelihood estimation and uniform inference with sporadic identification failure", *Journal of Econometrics*, vol. 173, 36–56, Supplemental Appendices.

Andrews, D.W.K. and Guggenberger, P. (2009): "Hybrid and size-corrected subsampling methods", *Econometrica*, vol. 77, 721–762.

Andrews, D.W.K. and Guggenberger, P. (2010): "Asymptotic size and a problem with subsampling and with the m out of n bootstrap", *Econometric Theory*, vol. 26, 426–468.

McCloskey, A. (2017): "Bonferroni-based size-correction for nonstandard testing problems", *Journal of Econometrics*, vol. 200, 17–35.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference: Inequality, Order and Shape Restrictions.* John Wiley & Sons, Hoboken, NJ.

White, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, vol. 48, 817–838.

Table 8: Rejection Frequencies for Different Values of γ and β_3

β_3	LR	CLR	LR-uniform
Null hypothesis, $\gamma = 0$			
0	0.1094	0.0497	0.0092
0.01	0.1159	0.0468	0.0143
0.05	0.1196	0.0497	0.0230
0.1	0.1150	0.0484	0.0251
0.25	0.1116	0.0456	0.0232

γ	LR	CLR	LR-uniform
Alternative hypotheses with $\beta_3 = 0$			
0	0.1083	0.0498	0.0119
0.01	0.6767	0.3112	0.2850
0.05	1.0000	0.9907	1.0000
0.1	1.0000	1.0000	1.0000
0.25	1.0000	1.0000	1.0000